UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DEVELOPMENT OF LC-MS/MS METHODS FOR THE ANALYSIS OF REACTIVE METABOLITE PROTEIN TARGETS

DISSERTATION
PRESENTED
AS PARTIAL FULFILLMENT
OF THE DOCTORATE IN CHEMISTRY

BY
MAKAN GOLIZEH

DECEMBER 2015

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

DÉVELOPPEMENT DE MÉTHODES LC-MS/MS POUR L'ANALYSE DE PROTÉINES CIBLES DES MÉTABOLITES RÉACTIFS

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN CHIMIE

PAR
MAKAN GOLIZEH

DÉCEMBRE 2015

ACKNOWLEDGEMENTS

This Ph.D. program was a unique opportunity for me to gain a deeper understanding of bio-analytical mass spectrometry while enjoying a continuous sense of satisfaction through constant learning, improvement and productivity. This was, of course, not possible without the support of a number of people and organizations that, in the past four years, have effectively contributed to the content and context of my research. Therefore, I wish to cordially thank, first and foremost, my supervisor, Professor Lekha Sleno, for accepting me as a graduate student, training me with patience and precision, giving me the opportunity to engage multiple research projects in an independent yet coached manner, and for supporting me in all the aspects of my professional and personal life; all the former and present members of Professor Sleno's team, particularly Dr. André LeBlanc, Leanne Ohlund, Yasmin Boukhedimi, Maxime Sansoucy, Biao Ji, and Ghazaleh Moghaddam for being excellent co-workers and unforgettable friends; our resourceful undergraduate interns, Julie-Anne Jauffrit, Christina Schneider and Lucie Huart, for their assistance; Professor Isabelle Marcotte and Professor Steve Bourgault, for patiently accompanying my research and providing helpful comments; Professor Karen Waldron for kindly accepting to read and evaluate this thesis; Professor Huu Van Tra for being a supportive program director; the staff of the Department of Chemistry for always being nice and cooperative, particularly Sonia Lachance, Marie-Josée Crevier, Odette Desrosiers, Sylvie Lemieux, Isabelle Rheault, Mathieu Maurin-Soucy and Jacqueline Hue Tieu; the Faculty of Science and UQÀM for providing a positive and constructive environment; *Groupe de Recherche Axé sur la Structure des Protéines* (GRASP) and *Fondation UQÀM*, for graduate student scholarships; *Ministère de l'éducation du Québec*, for additional financial support; and finally, my brother, Ashkan, for always being understanding, caring and supportive before and during my Ph.D. studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABC	ATP-binding cassette
ACN	acetonitrile
ADR	adverse drug reaction
APAP	*N*-acetyl-*p*-aminophenol (acetaminophen)
APCI	atmospheric-pressure chemical ionization
API	atmospheric-pressure ionization
APPI	atmospheric-pressure photoionization
ASK1	apoptosis signal-regulating kinase 1
ATP	adenosine triphosphate
AX	anion exchange
CATWAX	mixed-bed weak cation/weak anion exchange
CE	capillary electrophoresis
CI	chemical ionization
CID	collision-induced dissociation
CMC	critical micelle concentration
COX	cyclooxygenase
CV	coefficient of variation
CX	cation exchange
CYP	cytochrome P450
DA	differential analysis
DAB	*N*,*N*-dimethyl-4-aminoazobenzene
DBS	dynamic background subtraction
DDA	data-dependent acquisition
DDM	*n*-dodecyl β-D-maltoside
DENN	differentially expressed in neoplastic versus normal cells
DIA	data-independent acquisition

DILI	drug-induced liver injury
DMEM	Dulbecco's modified Eagle's medium
DOTA	1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid
dR	deoxyribose
DTT	dithiothreitol
ECD	electron-capture dissociation
ECM	extracellular matrix
EI	electron ionization
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
ERLIC	electrostatic repulsion-hydrophilic interaction chromatography
ESI	electrospray ionization
ETC	electron transport chain
ETD	electron-transfer dissociation
ExPASy	expert protein analysis system
FAD	flavin adenine dinucleotide
FADP	flavin adenine dinucleotide phosphate
FASTA	fast-all
FBS	fetal bovine serum
FC-14	*n*-tetradecyl phosphocholine
FDR	false discovery rate
FMO	flavin-containing monooxygenase
FT-ICR	Fourier-transform ion cyclotron resonance
GC	gas chromatography
GDP	guanosine diphosphate
GE	gel-electrophoresis
GELFrEE	gel-eluted liquid fraction entrapment electrophoresis
Gluc	glucose
GO	gene ontology, glyoxal

GPCR	G-protein coupled receptor
GRAVY	grand average of hydropathy
GSH	glutathione
GST	glutathione *S*-transferase
GTP	guanosine triphosphate
HILIC	hydrophilic interaction liquid chromatography
HLB	hydrophilic-lipophilic balanced
HLM	human liver microsomes
HLS	human liver S9 fraction
HNE	4-hydroxy-2-(*E*)-nonenal
HPLC	high-performance (pressure) liquid chromatography
HR	high-resolution
HSAB	hard and soft acids and bases
IAM	2-iodoacetamide
ICAT	isotope-coded affinity tag
IDA	information-dependent acquisition
IDE	integrated development environment
IEF	isoelectric focusing
IL-1β	interleukin-1beta
IMP	integral membrane protein
IP	immunoprecipitation
IRMPD	infrared multi-photon dissociation
iTRAQ	isobaric tags for relative and absolute quantitation
JNK	c-Jun *N*-terminal kinase
LC	liquid chromatography
MADD	MAPK–activating death domain protein
MALDI	matrix-assisted laser desorption ionization
MAPK	mitogen-activated protein kinase
MAX	mixed-mode anion exchange

MCX	mixed-mode cation exchange
MDA	malondialdehyde
MDLC	multidimensional liquid chromatography
MeCAT	metal-coded affinity tag
MGO	methylglyoxal
MGST1	microsomal glutathione *S*-transferase 1
MLM	mouse liver microsomes
MLS	mouse liver S9 fraction
MMP	matrix metalloproteinase
MPO	myeloperoxidase
MRM	multiple-reaction monitoring
MS	mass spectrometry
MSA	multiple sequence alignment
MS/MS	tandem mass spectrometry
MudPIT	multidimensional protein identification technology
MWCO	molecular weight cut-off
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NAPQI	*N*-acetyl-*p*-benzoquinone imine
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NOX	NADPH oxidase
NSAID	non-steroidal anti-inflammatory drug
NU5M	NADH-ubiquinone oxidoreductase chain 5
OA	osteoarthritis
ONE	4-oxo-2-(*E*)-nonenal
PANTHER	protein analysis through evolutionary relationships
PAPS	3'-phosphoadenosine-5'-phosphosulfate
PIM	percent identity matrix

PMF	peptide mass fingerprinting
PMP	peripheral membrane protein
POE	polyoxyethylene
PPI	protein-protein interaction
PSD	post-source decay
PS-DVB	polystyrene-divinylbenzene
PSM	peptide spectral matching
PTGS	prostaglandin G/H synthase
PTM	post-translational modification
PUFA	polyunsaturated fatty acid
QET	quasi-equilibrium theory
RA	rheumatoid arthritis
ReTOF	reflectron time-of-flight
RLM	rat liver microsomes
RLS	rat liver S9 fraction
RNS	reactive nitrogen species
ROS	reactive oxygen species
RP	reversed-phase
SCX	strong cation exchange
SDC	sodium deoxycholate
SDS	sodium *n*-dodecyl sulfate
SDS-PAGE	sodium *n*-dodecyl sulphate polyacrylamide gel-electrophoresis
SFC	supercritical fluid chromatography
SILAC	stable isotope-labeling with amino acids in cell culture
SIM	selected-ion monitoring
SOD	superoxidase dismutase
SPE	solid-phase extraction
SRM	selected-reaction monitoring
SULT	sulfotransferase

SWATH	sequential window-acquisition of all theoretical mass spectra
TCA	trichloroacetic acid
TDC	time-to-digital converter
TFA	trifluoroacetic acid
TIC	total ion chromatogram
TMD	transmembrane domain
TNF	tumor necrosis factor
TOF	time-of-flight
TPDB	target protein database
TRAF2	TNF receptor-associated factor 2
Tris	tris(hydroxymethyl)aminomethane
UDP	uridine diphosphate
UGT	UDP-glucuronyltransferase
UHPLC	ultra-high-pressure HPLC
UV	ultraviolet
WAX	weak anion exchange
WCX	weak cation exchange
XAO	xanthine oxidase
XIC	extracted ion chromatogram

LIST OF SYMBOLS AND UNITS

Å	Angstrom
B	magnetic sector
°C	degree centigrade (Celsius)
cc	cubic centimeter
cps	cycle per second
Da	Dalton
ΔE_0	potential difference
E	total energy
E_0	potential
E_{ex}	excess energy
E_{int}	internal energy
g	gram, gravity (g-force)
h	hour
η	chemical hardness
k_E	rate constant of a gas-phase unimolecular reaction
l	liter
M	molar
m	meter
μ	micron
min	minute
MW	molecular weight
m/z	mass-to-charge
N	number of electrons
n	number of analytical replicates
ν	number of vibrational states

NaN	not a number
P	peak capacity
pI	isoelectric point
ppm	part per million
psi	pounds per square inch
Q	quadrupole
q	radiofrequency-only quadrupole
rh	recombinant human
rpm	round per minute
RT	retention time
s	total degree of freedom
s	second
s^{-1}	per second
S/N	signal-to-noise
V	Volt
v/v	volume per volume
w/v	weight per volume
w/w	weight per weight
z	charge

RÉSUMÉ

Un grand nombre de médicaments, de produits naturels et de substances endogènes subissent les transformations métaboliques en formant des « métabolites réactifs » qui peuvent réagir avec des biomolécules, altérer leurs fonctions moléculaires et affecter des processus biologiques. Le foie joue un rôle prédominant dans le métabolisme, et les protéines hépatiques sont souvent ciblées par les métabolites réactifs causant potentiellement l'hépatotoxicité. L'identification de ces protéines est donc essentielle pour mieux comprendre les mécanismes reliés à cette toxicité. Plusieurs essais ont été effectués afin d'étudier les liaisons covalentes sur les protéines et identifier leurs cibles potentielles. Toutefois, la faible abondance des protéines modifiées, les problèmes techniques, et l'absence de méthodes appropriées pour analyser des échantillons biologiques ont été les défis limitant le progrès de cette recherche. Parmi les techniques analytiques employées en bioanalyse, la spectrométrie de masse (MS) est devenue la méthode de choix en raison de sa capacité exceptionnelle d'acquérir des grandes quantités de données qualitatives et quantitatives à partir de mélanges complexes. La spectrométrie de masse en tandem (MS/MS) couplée à la chromatographie liquide (LC) a permis l'analyse efficace des échantillons protéiques avec une grande sensibilité dans la recherche protéomique.

La thématique principale de cette thèse a été le développement d'une approche analytique pour identifier les protéines cibles des métabolites réactifs et leurs sites de modifications par LC-MS/MS. Cette recherche a été organisée en trois phases, selon un des principaux défis posés dans ce type d'analyse : (1) préparation d'échantillon, (2) détection par LC-MS/MS, et (3) traitement des données. Premièrement, la méthode analytique a été optimisée pour l'analyse protéomique dans les fractions S9 et microsomales de foie de rat, souris, et humain. Puis, des microsomes de rat ont été incubés avec l'acétaminophène, formant un métabolite réactif, afin de générer des adduits protéiques, puis analysés par la méthode développée en utilisant une stratégie de traitement de données multi-étage. Une approche plus ciblée a été également établie afin d'étudier la modification covalente de métalloprotéase matricielle 13 par le produit de peroxydation lipidique, 4-hydroxynonénal.

Mots-clés : liaison covalente, métabolites réactifs, LC-MS/MS, microsomes de foie de rat, acétaminophène, métalloprotéase matricielle 13, 4-hydroxynonénal

ABSTRACT

Numerous drugs, naturally-occurring products and endogenous compounds undergo metabolic transformations to form "reactive metabolites". These species can covalently bind to biomolecules, including proteins, altering their molecular functions affecting biological processes. The liver plays a predominant role in metabolic transformations and hepatic proteins are often targeted by reactive metabolites potentially leading to hepatotoxicity. Identification of these proteins is therefore important to understand the mechanisms involved in this type of toxicity.

Several previous attempts have been made to study protein covalent binding reactions and identify their potential targets. However, low abundance of the modified proteins, technological issues, and lack of appropriate methods for efficient analysis of complex biological samples are challenges that have limited the progress of this research. Amongst the analytical techniques employed in bioanalysis, mass spectrometry (MS) has become the method of choice due to its incredible capacity to acquire large amount of qualitative and quantitative information from complex mixtures. Tandem mass spectrometry (MS/MS) in combination with liquid chromatography (LC) has enabled efficient analysis of protein samples with high sensitivity fostering new possibilities in proteomics research.

The main theme of this dissertation was to systematically develop an analytical approach to identify reactive metabolite protein targets and their sites of modification using LC-MS/MS. This research was designed in three phases to address the main challenges posed in this type of analysis: (1) sample preparation, (2) LC-MS/MS detection, and (3) data analysis strategies. The analytical method was first optimized for proteomics analysis of rat, mouse and human liver S9 and microsomal fractions. Next, rat liver microsomes were incubated with acetaminophen, known to form a reactive metabolite, to produce protein adducts, which were then subjected to the developed method and identified using a multi-stage data processing workflow. A more targeted approach was also established to study covalent binding of matrix metalloproteinase 13 by the lipid peroxidation product, 4-hydroxynonenal.

Keywords: covalent binding, reactive metabolites, LC-MS/MS, rat liver microsomes, acetaminophen, matrix metalloproteinase 13, 4-hydroxynonenal

CHAPTER ONE

INTRODUCTION

Most compounds which a living organism is exposed to are changed upon excretion. The sophisticated molecular machinery of the cell encompasses a myriad of highly specialized mechanisms to cope with different groups of chemicals that bind to or cross its plasma membrane. These biotransformations however, are not necessarily flawless, and under certain conditions, can jeopardize cell homeostasis and survival.

Scientific and technological breakthroughs in the last decades gave rise to novel areas of research that enabled a clearer understanding of biological mechanisms. Bioanalytical chemistry, or bioanalysis, is a sub-discipline of analytical chemistry that concentrates on qualitative and quantitative analysis of small and large molecules in biological systems using efficient instrumental techniques such as chromatographic and electrophoretic separation, ligand binding assays, mass spectrometry (MS), and nuclear magnetic resonance (NMR). The expanding role of bioanalytical chemistry in academic and industrial environments has made it an important area of research for scientists in chemistry and biochemistry. The exceptional capacity of bioanalytical techniques in providing high-quality data on the chemical composition of complex matrices have, in recent years, facilitated numerous scientific endeavors that depend on accurate measurements in biological samples, such as pharmacological, molecular toxicology, forensic, and environmental research.

This introductory chapter will briefly discuss the chemistry of intrinsically reactive products of biotransformation, known as the "reactive metabolites", which are the major focus of this dissertation, followed by a short overview on bioanalytical techniques used in this work. The chapter will be concluded by discussing current challenges in this area and how they are addressed in this thesis.

1.1 Molecular Toxicology of Reactive Metabolites

Metabolism usually produces "non-toxic" metabolites, which are relatively polar and can readily be excreted from the living organism and thus detoxified. However, certain xenobiotics and endogenous compounds form highly reactive metabolites, which can interact with vital intracellular macromolecules, and result in toxicity. In the 1940s and 1950s, James and Elizabeth Miller demonstrated for the first time that *N*,*N*-dimethyl-4-aminoazobenzene (DAB), a hepatocarcinogen in rats, would covalently bind to proteins and nucleic acids through a process they called "metabolic activation". They also demonstrated that covalent binding of these species was an essential part of the carcinogenic process (Conney, 2001).

Normally, reactive metabolites can be detoxified via a series of enzymatically-assisted covalent reactions to endogenous scavenger molecules known as the "conjugation reactions". However, overabundance or excessive reactivity may help these chemicals skip or saturate designated conjugation mechanisms to trigger undesired reactions within the tissue depending on the nature of the reactive species and the physiology of the organism (Attia, 2010).

The following section briefly discusses major intracellular sources of the reactive metabolites, their formation and fate within the organism, the chemistry of the covalent binding reaction, and its preferred molecular targets.

1.1.1 Biologic Oxidation and Oxidative Stress

Oxygen is the vital element of aerobic life in which oxidation of organic compounds is exploited for the release of free energy. However, not only is O_2 fatal to anaerobic bacteria, due to its high chemical reactivity, excess oxygen is also toxic to aerobic organisms. Even at non-toxic concentrations, aerobic organisms must deal with the toxic consequences of partial oxidation-reduction reactions, that is the formation of reactive oxygen-derived compounds (Josephy, 1997).

An oxidation-reduction, or redox, reaction is of the general form

$$A_{red} + B_{ox} \longrightarrow A_{ox} + B_{red}$$

which can be written as two separate yet dependent and complementary half-reactions involving an electron-donor (reductant) and electron-acceptor (oxidant) species:

$$A_{red} \longrightarrow A_{ox} + n\,e^-$$

$$B_{ox} + n\,e^- \longrightarrow B_{red}$$

The direction of a redox reaction depends on which species has a greater tendency (or potential, E_0) to accept the available electrons. The potential difference (ΔE_0) of a redox system is an important thermodynamic factor, and determines if a redox reaction can proceed spontaneously.

In living organisms, redox reactions are catalyzed by "oxidoreductases", a large class of enzymes that utilize pyridine- or flavin-cofactors (Figure 1.1) to transfer electrons from one molecule to another. Oxidoreductases are generally classified into four groups (oxidases, dehydrogenases, hydroperoxidases, and oxygenases) and reduce or oxidize a wide range of organic functional groups and inorganic elements (Murray *et al.*, 2009).

Redox reactions are also the core of the electron transport chain (ETC), the essential component of photosynthesis and cellular respiration in prokaryotic and eukaryotic cells. In eukaryotes, the ETC transfers electrons from NADH to O_2 (ΔE_0 = 1.1 V) through three large protein complexes: NADH-ubiquinone oxidoreductase (Complex I), ubiquinone-cytochrome c oxidoreductase (Complex III), and cytochrome c oxidase (Complex IV). This electron transfer creates an electrochemical proton gradient across the inner mitochondrial membrane that drives adenosine triphosphate (ATP) synthesis (Murray *et al.*, 2009).

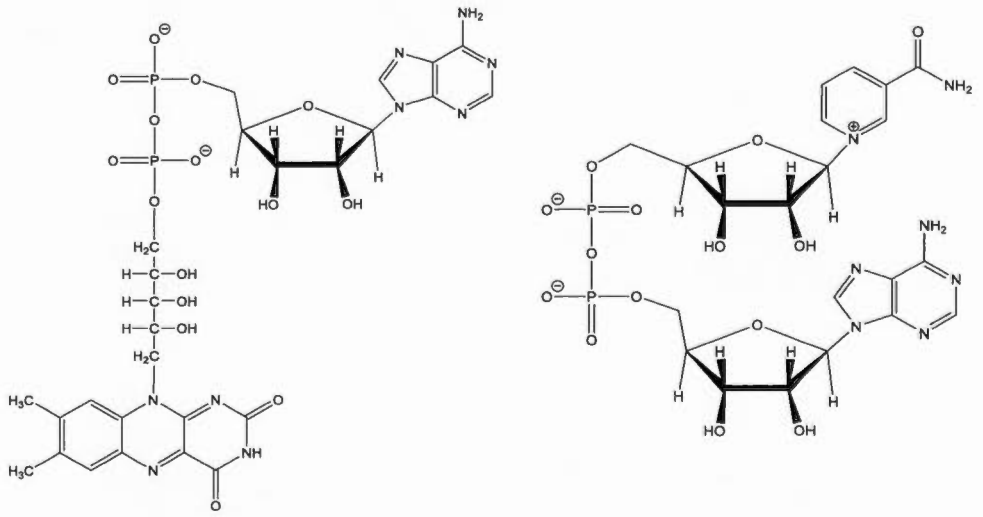


Figure 1.1 Chemical structure of flavin adenine dinucleotide (FAD) (left) and nicotinamide adenine dinucleotide (NAD) (right) cofactors involved in biologic redox reactions

The final stage of the ETC involves a four-step reduction of molecular oxygen (O_2) to water:

$$O_2 \xrightarrow[-0.33\ V]{e^-} O_2^{\bullet -} \xrightarrow[0.89\ V]{e^-} H_2O_2 \xrightarrow[0.38\ V]{e^-} \dot{O}H \xrightarrow[2.32\ V]{e^-} H_2O$$

The above reduction reaction has a very large positive ΔE_0 value and is therefore thermodynamically favorable in the presence of almost any biochemical reductant. In the cell, the transfer of four electrons to O_2 involves iron and copper ions, and is accomplished in a controlled and concerted manner; however, the release of intermediary products (superoxide, peroxide, hydroxyl radical) cannot be fully avoided (Josephy, 1997). Oxygen reduction intermediates are extremely reactive and immediately react with other species to give rise to radical and non-radical metabolites, including "reactive oxygen species (ROS)" and "reactive nitrogen species (RNS)", representing the major sources of oxidative and nitroxidative stress. A list of the most important reactive oxygen metabolites is given in Table 1.1 (Kohen and Nyska, 2002).

Table 1.1 Radical and non-radical oxygen metabolites

Radical	Non-radical
Oxygen bi-radical ($O_2^{\bullet\bullet}$)	Hydrogen peroxide (H_2O_2)
Superoxide ion ($O_2^{\bullet -}$)	Organic peroxide (ROOH)
Hydroxyl (HO•)	Hypochlorous acid (HOCl)
Peroxyl (ROO•)	Ozone (O_3)
Alkoxyl (RO•)	Aldehydes (RCOH)
Nitric oxide (NO•)	Singlet oxygen ($^1O_2$)
	Peroxynitrite (ONOOH)

Oxidative stress can be induced either directly by "primary ROS" or indirectly via partial oxidation of biomolecules, such as lipids and carbohydrates. The products of these reactions, which are usually classified under "secondary ROS", are also highly reactive towards biomolecules such as proteins and nucleic acids. Secondary

ROS include malondialdehyde (MDA), 4-hydroxy-2-(*E*)-nonenal (HNE), 4-oxo-2-(*E*)-nonenal (ONE), and 2-oxoaldehydes such as glyoxal (GO) and methylglyoxal (MGO). The chemistry of these reactions will be discussed in more details in Chapter 6. Free radical chain reactions, such as autoxidation reactions, metal ions (especially iron, copper, and cobalt), and certain oxidoreductases, including NADPH oxidase, xanthine oxidase, superoxide dismutase, and myeloperoxidase are other key players of ROS formation and oxidative stress. Figure 1.2 summarizes chemical species, enzymes and the mechanisms involved in these processes (Chondrogianni *et al.*, 2014).

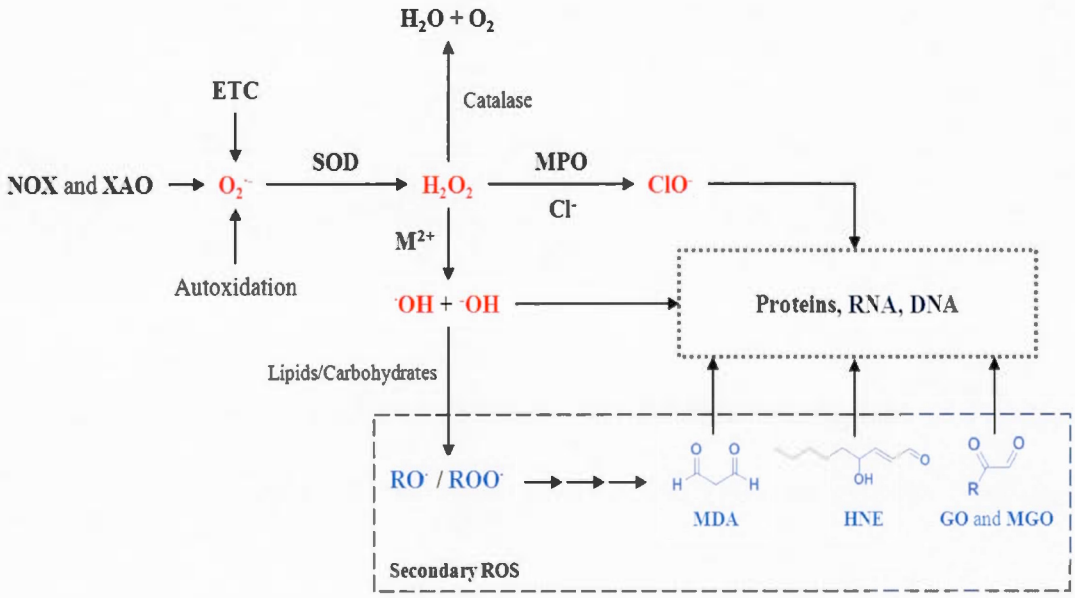


Figure 1.2 Key players of oxidative stress and potential molecular targets. ETC: electron transfer chain; GO: glyoxal; HNE: 4-hydroxy-2-(*E*)-nonenal; MDA: malondialdehyde; MGO: methylglyoxal; MPO: myeloperoxidase; NOX: NADPH oxidase; ROS: reactive oxygen species; SOD: Superoxidase dismutase; XAO: xanthine oxidase. M^{2+} is a bivalent metal ion.

Living organisms have adapted to continuously control and tightly regulate the presence of oxygen metabolites. This is extremely important for maintaining vital cellular and biochemical functions and any interference or imbalance leading to an uncontrolled state can cause excessive oxidative stress and cellular damage.

1.1.2 Xenobiotic Metabolism

The Unified Medical Language System classifies any chemical substance that is foreign to the biological system including natural products, drugs, environmental agents, *etc.* as a "xenobiotic".[1]

In animals, xenobiotics enter the body through digestion, inhalation, or intracutaneously, and exit via urine, faeces, sweat or breath. Most environmentally persistent xenobiotics are hydrophobic, thus can easily cross biological membranes. For the same reason, these compounds cannot be removed by ultrafiltration of the blood in the kidneys. Hydrophobic xenobiotics can sometimes be extracted by the liver and excreted into faeces through bile; yet the majority of these species tend to be passively readsorbed and thus cannot be effectively eliminated. This problem is solved by metabolic transformation of lipophilic xenobiotics into more water-soluble conjugates (Josephy, 1997).

One of the liver's main physiological functions is the metabolism of xenobiotics into hydrophilic metabolites through a series of oxidation, reduction, or hydrolysis (phase I) followed by conjugation reactions (phase II). Cytochrome P450 (CYP) enzymes play a predominant role in the phase I metabolism of a large variety of chemical compounds (Williams, 2006). CYPs represent a group of heme-containing proteins that bind two atoms of oxygen, to form a molecule of water

[1] The word "*xenos*" is the Greek equivalent for "stranger".

together with the production of a metabolite, which is generally more polar than the original substrate, through the following reaction:

$$RH + O_2 + NADPH + H^+ \xrightarrow{CYP} ROH + H_2O + NADP^+$$

CYPs often catalyze hydroxylation, dealkylation, oxidation, ring-opening, and reduction reactions and, based on the nature of the electron transfer proteins, are usually classified into mitochondrial, microsomal, and bacterial P450 systems (Hanukoglu, 1996). In human, more than 50 CYP isoforms have been isolated among which the CYP1, CYP2 and CYP3 families are known to be more involved in xenobiotic metabolism, although most CYPs also catalyze metabolic conversion of a variety of endogenous compounds such as fatty acids, steroids, vitamins, bile acids and hormones. In higher animals, CYPs are localized in nearly all organs, especially the liver, small intestine, skin, nasal epithelia, lung and kidney. However, the highest concentration has been found in the liver (300 pmol/mg microsomal protein) and the intestinal epithelia (20 pmol/mg microsomal protein) (Martignoni *et al.*, 2006). Table 1.2 and Table 1.3 summarize human CYPs based on their substrates (Ioannides, 2008).

Table 1.2 Human cytochrome P450 enzymes and their substrates

Substrates	CYP involved
Fatty Acids	4A11, 4B1, 4F12, 2J2
Eicosanoids	4F2, 4F3, 4F8, 5A1, 8A1
Vitamins	24A1, 26A1, 26B1, 27B1
Steroids	7A1, 7B1, 8B1, 11A1, 11B1, 11B2, 17A1, 19A1, 21A2, 27A1, 39A1, 46A1, 51A1
Xenobiotics	1A1, 1A2, 1B1, 2A6, 2A13, 2B6, 2C8, 2C9, 2C18, 2C19, 2D6, 2E1, 2F1, 3A4, 3A5, 3A7

Table 1.3 Selected human cytochrome P450 xenobiotic substrates

CYP	Substrates
CYP1A2	acetaminophen, antipyrine, caffeine, clozapine, clomipramine, imipramine, phenacetin, propranolol, tacrine, theophylline
CYP2A6	cotinine, coumarin, 2,6-dichlorobenzonitrile, fadrozole, indole, losigamone, nicotine, 4-nitroanisole, quinolone
CYP2B6	arteether, 7-benzyloxyresorufin, benzphetamine, bupropion, cinnarizine, deprenyl, 7-ethoxycoumarin, testosterone
CYP2C9	antipyrine, diclofenac, flurbiprofen, ibuprofen, mefenamic acid, *R*-naproxen, phenytoin, tienilic acid, tolbutamide, *S*-warfarin
CYP2C19	amitriptyline, citalopram, clomipramine, diazepam, imipramine, *S*-mephenytoin, mephobarbital, omeprazole, proguanil, propranolol
CYP2D6	amitriptyline, aprindine, clomipramine, codeine, debrisoquine, desipramine, *N*-desmethylclomipramine, dextromethorphan, dihydro-codeine, ethyl-morphine, encainide, fluoxetine, haloperidol, hydro-codone, imipramine, maprotiline, MDMA, metoprolol, mexiletine, nortriptyline, paroxetine, perphenazine, propafenone, propranolol, sparteine, thioridazine, tramadol, trimipramine, zuclopenthixol
CYP2E1	acetaminophen, benzene, chlorzoxazone, dapsone, dimethyl-nitrosamine, enflurane, ethanol, halothane, 4-nitrophenol, salicylic acid
CYP3A4	alprazolam, amiodarone, carbamazepine, clomipramine, cortisol, cyclosporin A, diazepam, diltiazem, ethynyloestradiol, erythromycin, felopidine, ifosfamide, imipramine, lidocaine, midazolam, nifedipine, omeprazole, proguanil, propafenone, quinidine, tamoxifen, terfenadine, testosterone, toremifene, triazolam, verapamil, vinblastine

Phase II biotransformations modify the structure of xenobiotics by conjugating endogenous metabolites, often to increase their water-solubility. Moreover, some of

the functional groups used in these reactions, such as the carboxylate terminal glucuronate conjugates, are recognized by specific organic anion active transporters in the liver and kidney, facilitating excretion into the bile and urine, respectively. The majority of Phase II conjugations occur via electrophilic addition of an "activated donor" to nucleophilic centers in xenobiotic substrates, including oxygen in hydroxyl and nitrogen in amine groups. Glutathione (GSH), on the contrary, reacts with electrophilic centers of the xenobiotic or its reactive metabolite. Table 1.4 lists the most common xenobiotic conjugation reactions, their associated enzymes and subcellular localizations (Josephy, 1997; Badenhorst *et al.*, 2013).

Table 1.4 Selected xenobiotic conjugation reactions

Reaction	Enzyme	Cellular Component
Glucuronidation	UDP-glucuronyltransferase	Microsomal
Sulfation	Sulfotransferase	Cytosolic, Microsomal
Methylation	Methyltransferase	Cytosolic, Microsomal
Acetylation	Acetyltransferase	Cytosolic
Glycine conjugation	Glycine *N*-acyltransferase	Cytosolic, Microsomal
Glutathione conjugation	Glutathione *S*-transferase	Cytosolic, Microsomal

Depending on the organ involved, xenobiotic metabolism can sometimes involve multiple biotransformation routes. Figure 1.3 shows the metabolic pathway for *N*-acetyl-*p*-aminophenol (APAP), also known as paracetamol or acetaminophen. As illustrated, a therapeutic dose of APAP is primarily metabolized in the liver via glucuronylation and sulfation and then excreted into bile or blood by ATP-binding cassette (ABC) transporters. However, a small portion of the drug undergoes a phase I transformation to *N*-acetyl-*p*-benzoquinone imine (NAPQI), a reactive metabolite, which can subsequently be conjugated with GSH, enzymatically degraded, and excreted via urine (Mazaleuskaya *et al.*, 2015).

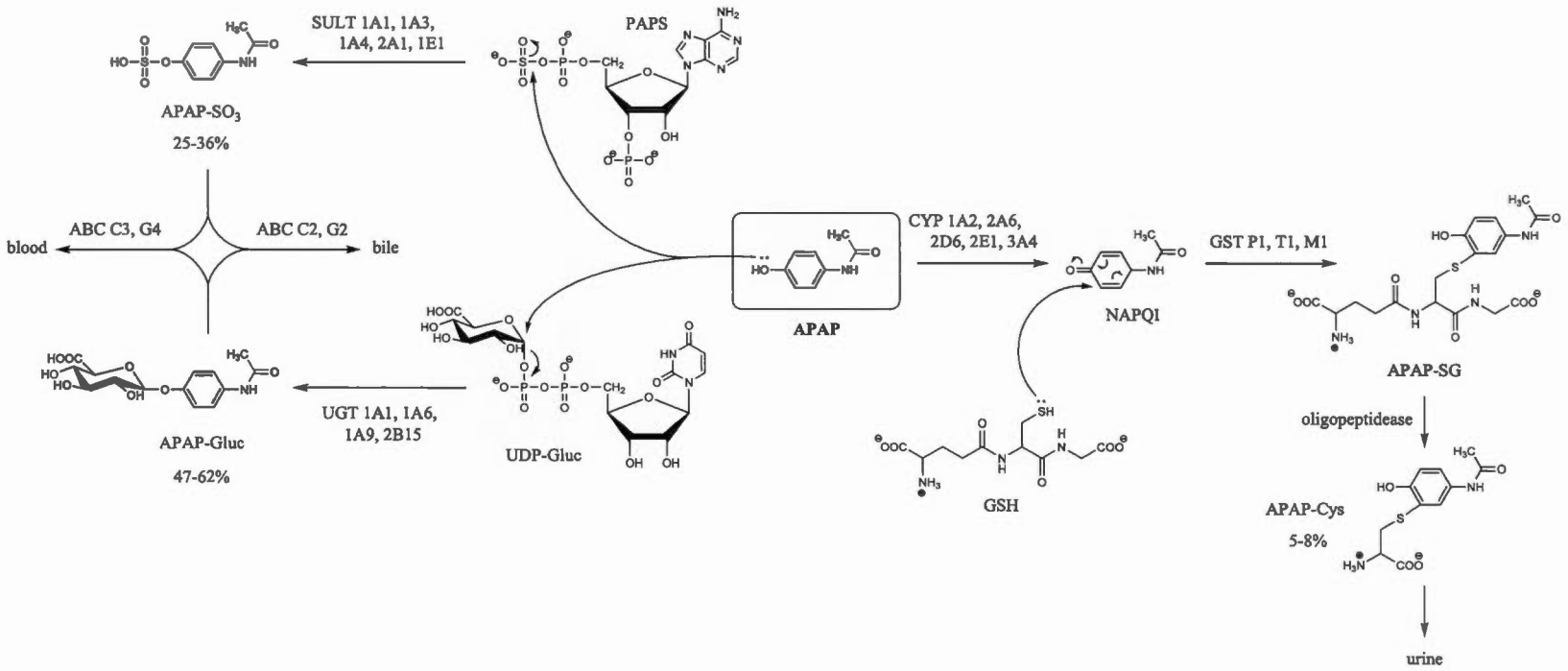


Figure 1.3 Summary of the acetaminophen metabolic pathway. The enzymes and intermediates involved as well as the excretion routes for each synthetic conjugate are also shown. ABC: ATP-binding cassette transporter; APAP: acetaminophen, CYP: cytochrome P450; Gluc: glucose; GSH: glutathione; GST: glutathione *S*-transferase; NAPQI: *N*-acetyl-*p*-benzoquinone imine; PAPS: 3'-phosphoadenosine-5'-phosphosulfate; SULT: sulfotransferase; UDP: uridine diphosphate; UGT: UDP-glucuronyltransferase.

Phase I and II biotransformations often serve as an efficient detoxification pathway; however, some CYP-catalyzed reactions may generate metabolites that are more reactive than the original xenobiotic ("bioactivation") and can covalently bind to endogenous compounds other than GSH or phase II activated donors. These reactions can particularly occur on the reactive centers of biological macromolecules, such as proteins and nucleic acids, affecting cellular processes, and potentially causing organ toxicity. The severity of the resulting insult largely depends on the extent of binding and the biochemical function of the target molecule(s). The subsequent pathological consequences will be a balance between the rate of tissue damage and the rate of cellular repair (Williams, 2006). The CYP-induced formation of NAPQI (Figure 1.3) is a well-studied example of reactive metabolite-associated toxicity since being first reported in 1973 by Jollow *et al.* APAP toxicity will be discussed in further detail in Chapter 5. Figure 1.4 illustrates the steps involved in the metabolism of xenobiotics and the relationship between reactive metabolites, covalent binding and toxicity.

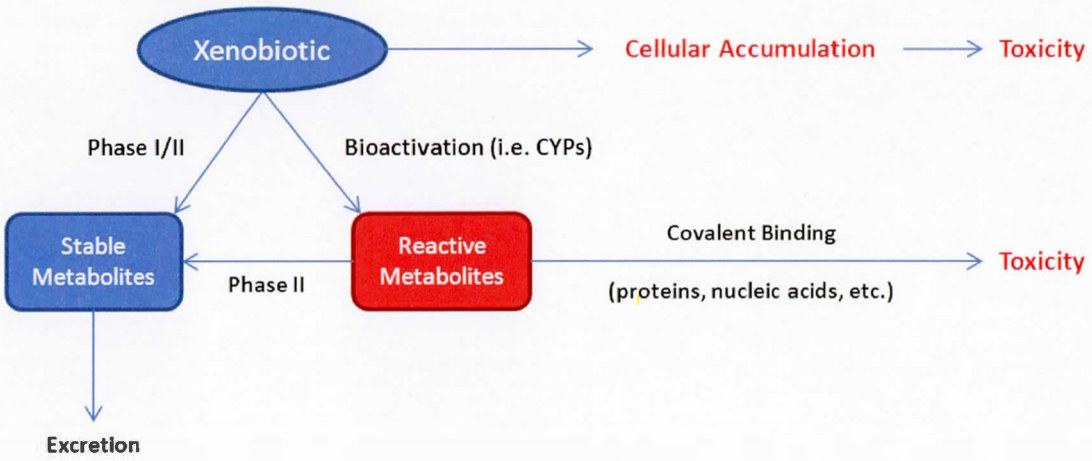


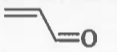
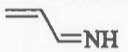
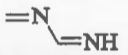
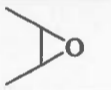
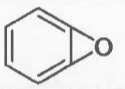
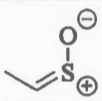
Figure 1.4 Relationship between xenobiotic metabolism and reactive metabolite-induced toxicity

Theoretically, both the nucleophilic and electrophilic reactive metabolites can undergo covalent binding reactions and induce toxicity; however, in nature, the reactions involving reactive electrophiles occur more frequently due to the lack of intrinsic highly electrophilic groups in biomolecules. Instead, nucleophilic centers are significantly present in proteins (cysteine, lysine, histidine, arginine, serine, threonine, selenocysteine, asparagine, aspartate and glutamate residues), nucleic acids (nitrogenous bases) and other compounds such as GSH. Therefore the study of reactive metabolites often involves the chemistry of biologically-occurring electrophiles (Gersch *et al.*, 2012).

Based on the mechanism of their reaction with nucleophiles, electrophilic metabolites can be classified into three major groups: (1) Michael acceptor systems, (2) ring-strained scaffolds, and (3) other electron-deficient systems. Table 1.5 lists the most common classes of reactive electrophiles under each category with selected xenobiotics known to form examples of each class (Zhou *et al.*, 2005; Kalgutkar, 2011).

According to the "electronic theory of acids and bases", electrophiles spontaneously react with nucleophiles, if energetic parameters are met. However, the selectivity of such a reaction mainly depends on the size and electron density of the species involved. From this perspective, an electrophile-nucleophile reaction is a Lewis acid-base reaction and its product results from the addition of two complete chemical entities, therefore often referred to as an "acid-base complex" or "adduct" (House, 2008). The selectivity of electrophilic (reactive) metabolites, the mechanisms of their reactions, and the types of adducts formed lie within an interesting research area and will be further discussed in the next section.

Table 1.5 Most common classes of electrophilic reactive metabolites

Reactive Electrophile	Structure	Example(s)
Michael Acceptor Systems		
α,β-Unsaturated carbonyl		carbamazepine, L-dopa, tolcapone
α,β-Unsaturated imine		acetaminophen, lidocaine
α,β-Unsaturated diimine		trimethoprim
Ring-Strained Scaffolds		
Epoxide		bergamottin, diallyl sulfone,
Arene oxide		capsaicin, imipramine, raloxifene
Other Electron-Deficient Systems		
Nitroso		erythromycin, phenacetin, sulfamethoxazole
Diazene		isoniazid
Iminium ion		aminopyrine, haloperidol, mirtazapine
Diazonium ion		hydralazine
Nitrenium ion		aristolochic acids, clozapine
Isocyanate	—N=C=O	sulofenur, tolbutamid, troglitazone
S-oxide		tienilic acid
Aldehyde/ketone		diclofenac, indomethacin
Acyl halides		chloramphenicol, halothane

1.1.3 Adduct Formation and Protein Targets

A covalent binding reaction on a protein does not automatically lead to toxicity. On the contrary, the efficacy of some drugs, such as penicillin, aspirin and omeprazole, relies on their ability to covalently bind to their protein targets (Williams, 2006). Nonetheless, irreversible chemical modification of a biomolecule by a reactive metabolite can affect its biological function and the cellular processes involving this molecule. It is therefore important to identify potential targets of a reactive metabolite to assess the extent and direction of such potential biological consequences.

Hard and soft acids and bases (HSAB) theory explains that hard electrophiles generally react with hard nucleophiles and soft species prefer the soft counterparts. In 1983, Parr and Pearson explained the concept of "chemical hardness" (η) as the second derivative of the total energy of a chemical system (E) with respect to changes in the number of electrons (N) at a fixed nuclear environment (z).

$$\eta = \frac{1}{2}\left(\frac{\partial^2 E}{\partial N^2}\right)_z$$

By this definition, "hard" species are small, have a high electron density, and are weakly polarizable, whereas "soft" species are large, have a low electron density, and can be readily polarized.

Applying HSAB theory to reactive metabolites, hard electrophiles, including cations, epoxides, and aldehydes, should react with hard nucleophilic centers such as the amine functional groups in DNA/RNA nucleobases and lysine residues in proteins. Soft electrophiles, such as Michael acceptor systems, would prefer soft nucleophiles, which include cysteine residues in proteins and glutathione. Free radicals can also react with lipids to initiate lipid peroxidation chain reactions. It is

however important to consider that, in addition to the chemical hardness, the microenvironment (pK_a, hydrophobicity, etc.) of a nucleophilic center in the tertiary structure of the target molecule can also affect an adduction reaction (Park *et al.*, 2005).

As discussed earlier, toxicological response of a covalent binding reaction largely depends on the biological role of the target molecule and the cellular processes affected by the target's structural alteration. These responses usually include (1) mutagenicity, teratogenicity or carcinogenicity, (2) necrotic or apoptotic cell death, and (3) hypersensitivity reactions (Williams, 2006). For example, the endogenous lipid peroxidation product, malondialdehyde (MDA), is known to covalently bind to guanine, adenine, and cytosine residues in DNA structure demonstrating mutagenic properties in bacterial and mammalian cells and carcinogenicity in rats (Figure 1.5) (Marnett, 2000). Phase I transformation of 5-aminosalicylic acid (also known as mesalazine), a non-steroidal anti-inflammatory drug (NSAID), by myeloperoxidase generates a quinone intermediate, which then binds to hemoglobin leading to hypersensitivity (Liu *et al.*, 1995). Similarly, the metabolism of several other drugs such as dapsone, carbamazepine, trimethoprim, and olanzapine results in covalent modification of neutrophil proteins leading to agranulocytosis (Zhou *et al.*, 2005).

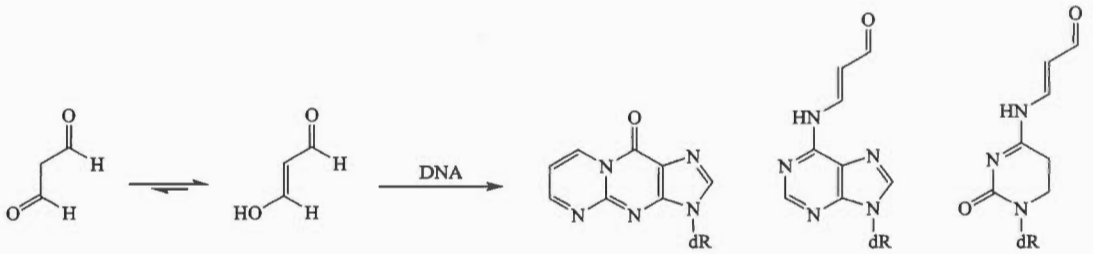


Figure 1.5 Covalent binding of malondialdehyde (MDA) with DNA leading to the formation of potentially mutagenic (modified) guanine (M_1G), adenine (M_1A), and cytosine (M_1C). dR: deoxyribose.

A comprehensive list of reactive metabolite *in vivo* target proteins can be found on Target Protein Database (TPDB) via http://tpdb.medchem.ku.edu:8080/protein_database (Hanzlik *et al.*, 2007). Figure 1.6 illustrates liver metabolism of trimethoprim, a bacteriostatic antibiotic, and the resulting covalent binding reaction of hepatic proteins (Lai *et al.*, 1999).

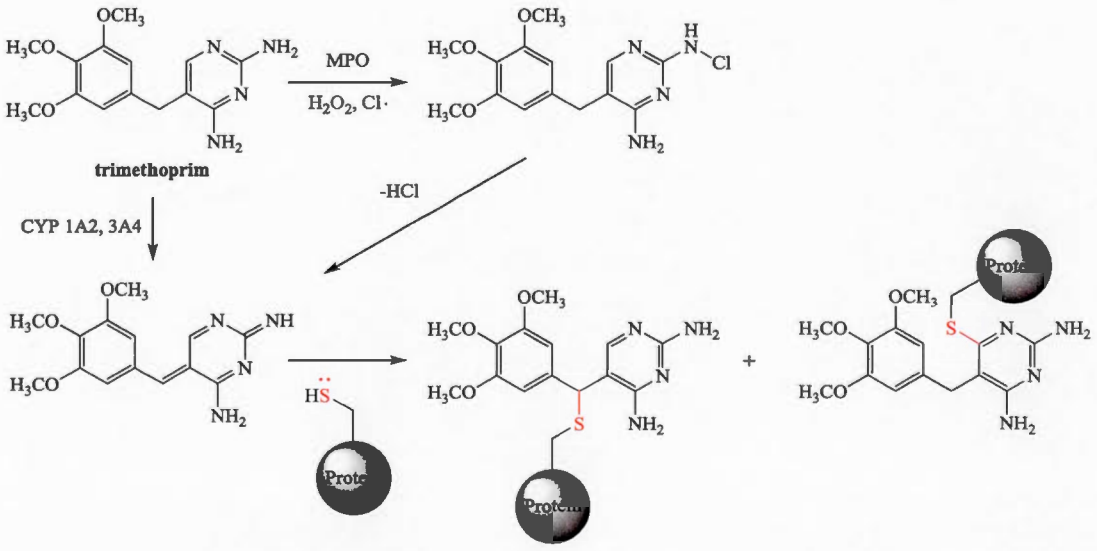


Figure 1.6 Liver metabolism of trimethoprim, formation of a reactive iminoquinone methide (an α,β-unsaturated diimine) metabolite, and its subsequent covalent binding to hepatic proteins. CYP: cytochrome P450; MPO: myeloperoxidase.

Approximately 1000 xenobiotics are known to induce hepatotoxicity (Chen *et al.*, 2014). More than half of the TPDB entries are liver protein targets. The liver plays a key role in the metabolism of xenobiotics and thus is exposed to high concentrations of reactive metabolites, making it an important target for covalent binding reactions. Although these reactions do not always lead to hepatotoxicity, drug-induced liver injury (DILI) is a serious challenge in drug development and often

results in regulatory actions including denied approval, use restrictions, or post-market withdrawal. It is also believed that several adverse drug reactions (ADRs) are also related to drug-induced organ toxicity (Williams, 2006). Identifying potential covalent target proteins can shed light on the biological mechanisms involved in a drug's ADRs and thus has become an important area of research in pharmacology and molecular toxicology. The following section will briefly discuss the methods developed in recent decades for systematic study of proteins in biological samples including techniques particularly adapted for covalent binding investigations.

1.2 Experimental Techniques: MS-Based Proteomics

The word "protein" was coined by Swedish chemist Jöns Jacob Berzelius, and introduced as a new class of organic substances by Dutch chemist Gerardus Johannes Mulder, who carried out the first elemental analysis of proteins in 1838 (Hartley, 1951). In 1845, Justus von Liebig, the German organic chemist, found that proteins, regardless of their slight difference in solubility under different conditions, are largely similar in their chemistry, and play an important role in biological processes in the body (Carpenter, 1986).

Analysis of proteins in biological systems has always been a challenge. Heat and pH sensitivity, extremely large molecular weights, and relatively high water-solubility make them difficult targets for traditional analytical techniques. Pioneering work on separation methods by Frederick Sanger paved the road for the development of protein and DNA/RNA sequencing, eventually bringing him two Nobel prizes. Tedious and time-consuming protein sequencing was later greatly improved by the development of phenylisothiocyanate sequencing chemistry by Edman in 1949, and its automation in 1967 (James, 1997).

The proteomics era began with the invention of sodium dodecyl sulphate polyacrylamide gel-electrophoresis (SDS-PAGE) (Kenrick and Margolis, 1970), and two-dimensional gel-electrophoresis (2D-GE) (O'Farrell, 1975; Klose, 1975) in the 1970s. 2D techniques significantly enhanced protein analysis resolving power and sensitivity, and enabled electro-blotting for protein identification with antibodies or Edman sequencing. However, low rate of detection of membrane proteins due to their poor water-solubility, limited reproducibility, and problems with spot identification in complex mixtures were major drawbacks of initial gel-based methods (James, 1997).

The development of bio-analytical mass spectrometry and the parallel growth of protein databases, in late 1990s and early 2000s (Aebersold and Mann, 2003; Washburn *et al.*, 2001), greatly enhanced the performance and throughput of traditional proteomics methods. The following section will briefly describe the state-of-the-art proteomics technology and use of mass spectrometry to address numerous challenges in protein analysis in biological samples.

1.2.1 Proteomics Technology

Proteomics is the large-scale study of proteins to obtain a global, integrated view of cellular processes at the protein level (Blackstock and Weir, 1999). The term "proteomics" was first introduced by Peter James (1997) from the Swiss Federal Institute of Technology to make an analogy with "genomics", the large-scale study of genes. The aim of a proteomics study was, at first, limited to describing protein expression changes under biological perturbations such as a specific disease state or drug treatment (Anderson and Anderson, 1998). Today, boundaries of proteomics have largely expanded to include qualitative and quantitative analysis of proteins, their functions, structure, and interactions with other molecules in a cell, within a tissue or a whole organism, or even across different species.

From a "systems biology" perspective, while genomics deals with the analysis of the complete set of DNA within an organism, and transcriptomics can be defined as the study of all RNA molecules and their relationships with protein biosynthesis, proteomics is the intermediary that links these areas to metabolomics, being the study of cellular processes involving small molecules known as metabolites (Mann *et al.*, 2013). The relationship between different levels of "omics" research and the type of molecules normally involved in each area of study are depicted in Figure 1.7.

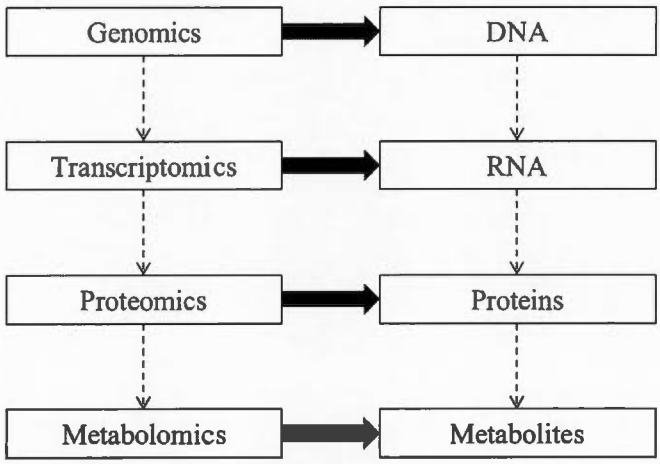


Figure 1.7 Simplified depiction of the four major domains of "omics" research and their primary molecular classes of interest

Initial proteomics analyses utilized gel-electrophoresis for separation and identification of proteins using manual (Sanger) sequencing. Modern methods however are usually based on automated gel-free approaches. Gel-based analysis includes protein separation followed by an in-gel protein digestion and subsequent peptide analysis using chemical sequencing techniques, such as Edman degradation (Edman *et al.*, 1950), or peptide mass fingerprinting (PMF) using matrix-assisted laser desorption ionization (MALDI) mass spectrometry. The latter is sometimes

referred to as gel-electrophoresis-mass spectrometry or GE-MS (Michaud and Snyder, 2002). Gel-based techniques are known for their accuracy and simplicity however, gel manipulation steps are time-consuming, and the method is not applicable to all proteins (Van Summeren *et al.*, 2012).

Gel-free proteomics usually involves in-solution digestion of the proteins prior to peptide fractionation or separation using one or multi-dimensional liquid chromatography (LC) coupled with atmospheric-pressure ionization (API) tandem mass spectrometry (MS/MS). Unlike gel-based techniques, gel-free proteomics could be readily used for high-throughput analysis and is applicable to most proteins; however, LC-MS/MS instruments are costly and require extensive method development (Van Summeren *et al.*, 2012). Figure 1.8 exhibits major steps involved in gel-based and gel-free proteomics analytical workflows.

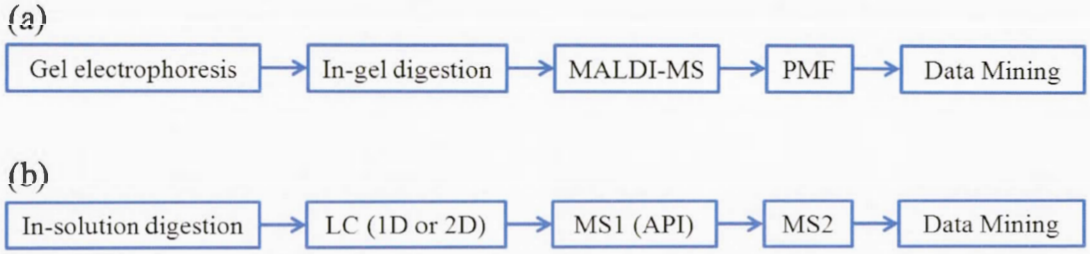


Figure 1.8 Different steps involved in a typical experimental workflow in a gel-based (a) or gel-free (b) proteomics approach. API: atmospheric-pressure ionization; LC: liquid chromatography; MALDI: matrix-assisted laser desorption ionization; MS: mass spectrometry; PMF: peptide mass fingerprinting.

A proteomics experiment can be "top-down" or "bottom-up" (Walther and Mann, 2010). Characterization of proteins by their molecular weight is known as "top-down" proteomics (McLafferty *et al.*, 2007). However, since the mass difference

between proteins with similar compositions is small, this method is not well suited for the analysis of protein mixtures, and thus "top-down" experiments are often limited to single protein analysis in relatively pure samples. For most experiments, proteins are first digested with resulting peptides separated and analyzed using tandem mass spectrometry. Computerized peptide sequencing algorithms then determine "peptide spectral matches" based on mass-to-charge (*m/z*) ratios and MS/MS fragmentation patterns. The software then compares putative peptide sequences to a protein database for identification. Since, in this approach, the analyte is qualitatively characterized based on the data obtained from its building blocks, this approach is often referred to as "bottom-up" or "shotgun" proteomics.

The bottom-up approach is known for its high compatibility with complex samples and is often used for quantitative and high-throughput studies. However, since proteins need to be denatured and digested prior to analysis, a portion of the structural information may be lost using this method (Van Summeren *et al.*, 2012). Typical bottom-up and top-down proteomics analysis workflows are demonstrated in Figure 1.9.

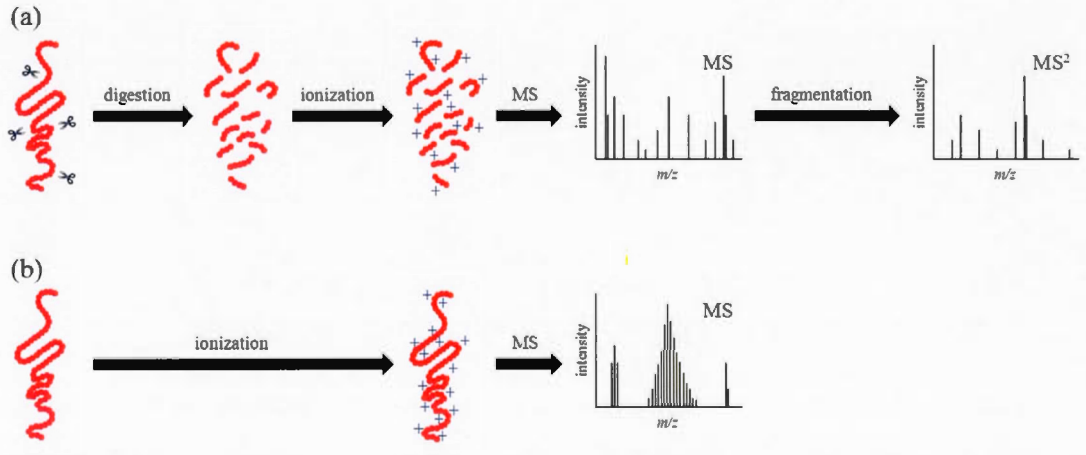


Figure 1.9 Steps involved in bottom-up (a) and top-down (b) proteomics analysis

In bottom-up proteomics, a peptide is usually fragmented at its amide (peptide) bonds, by collision with an inert gas such as He or N_2 at low pressure. This method of fragmentation is called collision-induced dissociation (CID). The resulting MS/MS spectrum is basically a list of *m/z* values of different fragments with some of the differences corresponding to the exact mass of an amino acid residue. Connecting the fragments with increasing size from the *N*-terminus (*b*-ion series) or *C*-terminus (*y*-ion series), in addition to the exact mass of the precursor ion, allows for the deduction of the peptide sequence from the series of residue-specific mass differences. This technique is known as *de novo* peptide sequencing. A CID-fragmented peptide (in positive ionization mode) and the structure of the *N*-terminal *a*, *b*, *c* and the *C*-terminal *x*, *y*, *z* fragment ions are depicted in Figure 1.10.

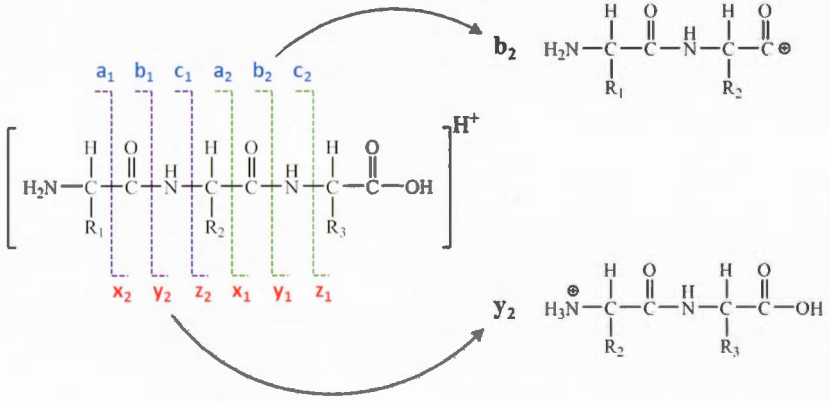


Figure 1.10 Peptide fragmentation via CID and structure of *b*- and *y*-type product ions often used for *de novo* peptide sequencing or peptide spectral matching

Proteomics analysis can be used for qualitative or quantitative purposes. Qualitative proteomics concentrates on protein identification and structural analysis including morphology, post-translational modification (PTM), and protein-protein

interaction (PPI) analyses. The state-of-the-art protein identification primarily relies on *de novo* peptide sequencing. Several probability-based searching sequence databases are developed for protein identification using mass spectrometry data including MASCOT (Perkins *et al.*, 1999), SEQUEST (Eng *et al.*, 1994), X!Tandem (Craig and Beavis, 2004) and Paragon (Shilov *et al.*, 2007). Qualitative proteomics information are routinely collected, reviewed and catalogued by user-based or curated proteomics knowledgebases such as the UniProt consortium (http://www.uniprot.org) and NCBI [2] (http://www.ncbi.nlm.nih.gov). Other web-based tools such as PANTHER [3] (http://www.pantherdb.org) (Thomas *et al.*, 2003), InnateDB (http://www.innatedb.ca) (Lynn *et al.*, 2008), and ExPASy[4] (http://www.expasy.org) (Artimo *et al.*, 2012) also provide possibilities for gene ontology (GO) classification, PPI analysis, and bioinformatics calculations, respectively.

The goal of quantitative proteomics is to determine the relative abundance of proteins between two experimental conditions or cellular states. Most quantitative proteomics approaches that are frequently used fall under one of the following two categories: (1) isotopic labeling and (2) label-free quantitation (Walther and Mann, 2010). In isotopic labeling, stable isotope-containing tags or amino acids are incorporated into the protein (or peptide). The sample containing the labeled protein ("heavy") will then be mixed with the non-treated ("light") sample. The MS signal ratio between the heavy and light samples enables relative quantitation at both protein and peptide levels. The most common techniques in use for stable isotope-labeling are SILAC[5] (Ong *et al.*, 2002), ICAT[6] (Gygi *et al.*, 1999) and iTRAQ[7] (Ross *et al.*, 2004). Figure 1.11 demonstrates a typical SILAC quantitation workflow where one cell population (light) is cultured in a growth medium containing regular amino acids

---

[2] National Center for Biotechnology Information
[3] Protein Analysis through Evolutionary Relationships
[4] Expert Protein Analysis System
[5] Stable Isotope-Labeling with Amino Acids in Cell Culture
[6] Isotope-Coded Affinity Tag
[7] Isobaric Tags for Relative and Absolute Quantitation

while the other (heavy) grows in a medium that only contains ^{15}N- or ^{13}C-labeled arginine or lysine. Label-free quantitation works by aligning separate LC-MS/MS runs of peptide mixtures to calculate differences in peak intensities of the same peptides detected in each run (Old *et al.*, 2005). In general, isotope-labeling techniques are time-consuming and costly and are not applicable to all proteins but are also more accurate than label-free methods which are, in turn, faster, cheaper and simpler (Walther and Mann, 2010).

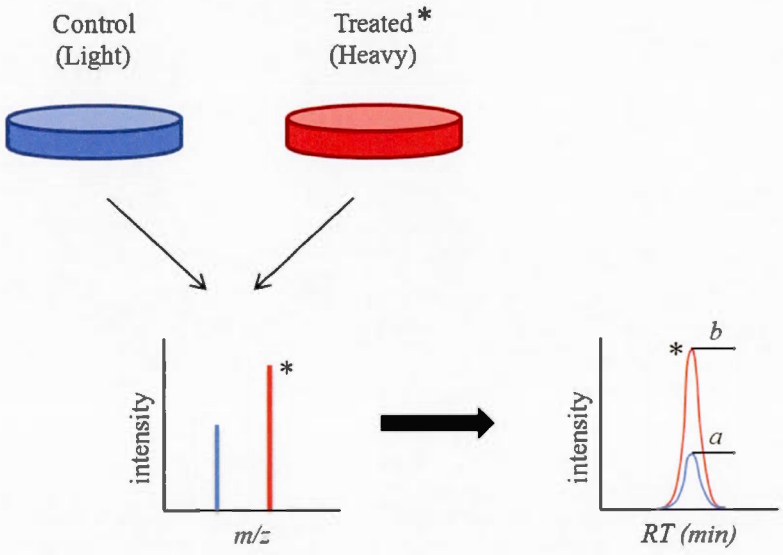


Figure 1.11 A typical SILAC quantitation workflow. The heavy-to-light LC-MS peak area ratio (*b/a*) in the [treated + control] mixture determines the relative abundance of labeled protein(s) in the treated sample.

MS is currently the method of choice in proteomics analysis (Altelaar and Heck, 2012). Advances in MS have been crucial for the development of proteomics and significantly contributed to the ability to identify and quantify low abundant proteins in complex mixtures. Several MS technologies have been developed, or combined with other analytical techniques, in recent years, to improve the resolving

power, sensitivity, and selectivity of proteomics analysis. High-resolution (Makarov, 2000), data-dependent (Mann *et al.*, 2001) and data-independent (Bern *et al.*, 2010; Gillet *et al.*, 2012) MS, in parallel with multidimensional (Washburn *et al.*, 2001), nano-scale (Tao *et al.*, 2011), and ultra-high-pressure liquid chromatography (UHPLC) (De Vos *et al.*, 2012) made drastic changes to the analytical capacity of modern proteomics. These techniques will be described later in this chapter.

1.2.2 LC-MS/MS

Mass spectrometry was first introduced by British physicist Joseph John Thomson, in 1897, with a series of experiments designed to study the nature of electric discharge in a high-vacuum cathode-ray tube coupled to a fluorescent screen (a "mass spectrograph"), leading to his Nobel Prize-winning discovery of the "electron". Early MS was often used to measure the atomic mass of elements and study their natural isotopes. Magnetic sector mass spectrometers, initially known as "calutrons", were devised by Canadian physicist Arthur Jeffrey Dempster, and used in the Manhattan Project for uranium enrichment during the Second World War. From the 1950s to the present, MS has changed dramatically, and numerous technological enhancements have turned it into one of the most powerful analytical tools available in chemistry, biochemistry, pharmacy, and medicine (Gross, 2004).

The basic principle of MS is to generate ions from organic or inorganic compounds using an "ionization" method, and then separate them by their mass-to-charge (*m/z*) ratio directly (quadrupole, ion trap) or indirectly based on a measurable physical property such as momentum (magnetic sector), kinetic energy (time-of-flight), or resonance frequency (orbital ion trap, ion cyclotron resonance). The compound may be ionized thermally or by electric fields, impacting energetic electrons, ions, or photons (Gross, 2004).

A mass spectrometer is comprised of three main components: (1) an ion source, (2) a mass filter or analyzer, and (3) a detector. A sample introduction system, such as a GC (gas chromatograph) or HPLC (high-pressure liquid chromatograph), vacuum supply, and data system are other parts of the MS system. The general scheme of a typical mass spectrometer is shown in Figure 1.12.

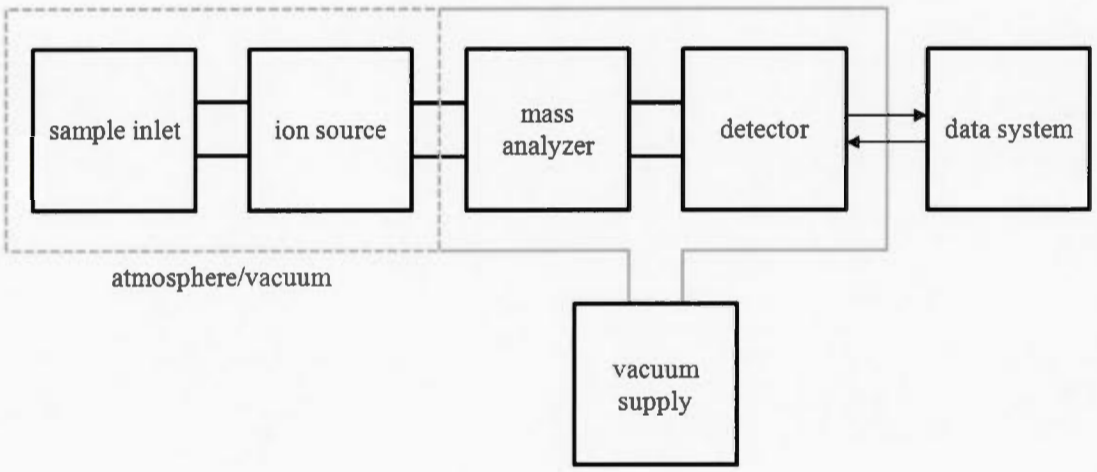


Figure 1.12 Major components of a typical mass spectrometer. Depending on system configuration, the sample inlet and ion source may or may not be under vacuum.

An ion source applies energy to the analyte (and matrix/medium) molecules to transform them into gas phase ions. This can occur under vacuum or at atmospheric pressure. Depending on the ionization process, an ion source can function in positive or negative ionization mode. Currently, the most frequently employed ion sources include electron ionization (EI), chemical ionization (CI), matrix-assisted laser desorption ionization (MALDI), and electrospray ionization (ESI). The energy applied to the sample to generate ions can transfer "excess energy" that thermodynamically destabilizes covalent bonds and breaks the molecule into "fragment" ions ("hard" ionization). EI is a "hard" ion source and is often coupled

with GC for analysis of volatile compounds and has a wide range of applications in environmental, forensic, and pharmaceutical analysis. MALDI and ESI are "soft" ion sources and thus are preferred for the analysis of biomolecules, which are fragile under harsher conditions, and thus are the ionization methods of choice in bio-analytical MS.

ESI, APCI (atmospheric-pressure chemical ionization) and APPI (atmospheric-pressure photoionization) do not require vacuum to help ionize the sample, and are known as API (atmospheric-pressure ionization) methods. API was the first technique to allow direct introduction of a solution phase enabling a more efficient LC-MS coupling compared to the previous less feasible counterparts, including the "moving belt" and PBI (particle beam interface) (Gross, 2004). ESI, the most common API method, was developed by American chemist John Bennett Fenn, laureate of the Nobel Prize of Chemistry in 2002.

There exists a great variety of mass analyzers as well. The separation of ions based on their *m/z* ratios occurs as a result of static or dynamic electric and magnetic fields (alone or combined). Most mass analyzers differ in the manner in which such fields are used to achieve separation. Analyzers can generally be divided into two major classes on the basis of their separation properties. Scanning analyzers, such as magnetic sector (B) and quadrupole (Q), transmit the ions with different *m/z* ratios successively along a time scale. Other analyzers allow the simultaneous transmission of all ions. Time-of-flight (TOF), Fourier-transform ion cyclotron resonance (FT-ICR) or orbital ion traps belong to this class. Mass analyzers can be coupled in tandem to enable fragmentation of precursor ions into fragment (product) ions.

Tandem MS (MS/MS) is an incredibly powerful technique for molecular structure determination and, as discussed earlier, the basis of MS-assisted peptide sequencing. Depending on the nature of the analyzer, precursor ions can be "activated" (obtain excess energy) through different methods to undergo selected

types of fragmentation. Ion activation can occur as a result of collision (collision-induced dissociation, CID), electron capture (electron-capture dissociation, ECD), electron transfer (electron-transfer dissociation, ETD), and infrared irradiation (infrared multi-photon dissociation, IRMPD), to name a few. Table 1.6 lists the most commonly used mass analyzers and their performance characteristics (Gross, 2004; de Hoffmann and Stroobant, 2007).

Tandem MS can be accomplished in a spatial or temporal domain. The physical coupling of mass analyzers constitutes the spatial domain and is often called "tandem-in-space" mass spectrometry. However, ion trapping mass analyzers including quadrupole ion trap, FT-ICR, and Orbitrap can isolate a precursor ion of a specific *m/z* value during one time interval, activate the ion, and collect (or store) product ions during a second time interval to obtain a product-ion mass spectrum by scanning the product ions out of the trap during a third time interval. This is an example of MS/MS in a temporal domain, and often referred to as "tandem-in-time" mass spectrometry. Tandem-in-time MS can be repeated multiple times and MS^n product ions can be scanned for as long as the signal-to-noise (S/N) ratio permits (Watson and Sparkman, 2007).

Table 1.6 Comparison of the most common mass analyzers

	Q	Ion trap	TOF	ReTOF	B	FT-ICR	Orbitrap
***m/z* limit**	4000	6000	> 1000000	10000	20000	30000	50000
Resolution (at *m/z* 1000)	2000	4000	5000	20000	100000	500000	100000
Mass accuracy (ppm)	100	100	200	10	< 10	< 5	< 5
Ion sampling	Continuous	Pulsed	Pulsed	Pulsed	Continuous	Pulsed	Pulsed
Pressure (Torr)	10^{-5}	10^{-3}	10^{-6}	10^{-6}	10^{-6}	10^{-10}	10^{-10}
Tandem MS:							
• **Configuration**	QQQ	-	-	TOF-TOF	B-B	-	-
• **Mode**	MS/MS	MS^n	-	MS/MS	MS/MS	MS^n	-
• **Collision energy**	Low	Low	-	Low/High	High	Low	-
• **Ion activation**	CID	CID	-	PSD/CID	CID	CID/ECD/IRMPD	CID/ETD

B: magnetic sector; CID: collision-induced dissociation; ECD: electron-capture dissociation; ETD: electron-transfer dissociation; FT-ICR: Fourier transform ion cyclotron resonance; IRMPD: infrared multi-photon dissociation; PSD: post-source decay; Q: quadrupole; ReTOF: reflectron time-of-flight; TOF: time-of-flight

In a product-ion scan, product ions produced by the decomposition of a given precursor are recorded by setting the first mass analyzer to a fixed value while a second mass analyzer acquires ion current over a range of *m/z* values. Product-ion spectra are rich sources of structural information as they include *m/z* values of the ions formed from the fragmentation of an activated precursor ion, usually as a result of collision with atoms or molecules of a neutral gas, such as He and N_2. The precursor ion is selected and directed into a "collision cell", where it collides with the collision gas converting some of its translation energy into "internal energy" (E_{int}). The difference between E_{int} and the amount of energy required to initiate decomposition (activation energy), known as the "excess energy" (E_{ex}), determines the type of fragmentation induced. According to "quasi-equilibrium theory (QET)", the rate constant of a gas-phase unimolecular reaction (k_E) can be defined as

$$k_E = \nu \left(\frac{E_{ex}}{E_{int}} \right)^{s-1}$$

where ν and s represent the number of vibrational states available to the activated molecule (the frequency factor) and its total degree of freedom, respectively. Generally, activated ions decomposing at rates below 10^5 s^{-1} will reach the detector without fragmentation ("stable ions"), whereas ions dissociating at above 10^6 s^{-1} fragment into product ions before reaching the detector ("unstable ions") (Gross, 2004).

Decomposition of an unstable ion can lead to the formation of ions, radicals and molecules (or "neutrals"). A "neutral loss" can provide useful structural information, particularly for peptide characterization. For example, loss of a molecule of water can usually be linked to the presence of Ser, Thr, and Glu residues, while Arg, Lys, Gln and Asn-containing peptides may lose molecules of ammonia. Loss of a •NO radical can be an indication of post-translational nitrosylation. In Chapter 6, neutral loss of 156.1150 Da was related to the modification by 4-hydroxynonenal.

Another useful MS/MS experiment, selected-reaction monitoring (SRM), or multiple-reaction monitoring (MRM), monitors the dissociation of one or a set of specific precursor ions (*m/z* 1) to one or a set of given product ions (*m/z* 2). Each [*m/z* 1, *m/z* 2] pair is called a "transition" or "reaction". MRM provides a better selectivity compared to "selected-ion monitoring" (SIM), in which only the *m/z* of the precursor is selected, and thus the former usually has better S/N making MRM an ideal technique for quantitative analysis in complex samples (Watson and Sparkman, 2007).

Product-ion scan and MRM are most widely used in MS-based proteomics for structure elucidation and selective detection of target peptides, respectively. A peptide's product-ion spectrum comprises *C*-terminal (*x*, *y*, *z*) and *N*-terminal (*a*, *b*, *c*) fragment ions allowing for *de novo* peptide sequencing. Other "diagnostic" fragment ions can provide useful information on PTMs and metabolically-occurring covalent modifications. MRM however is often employed for quantitative analysis where selectivity and sensitivity are crucial. MRM is utilized in combination with online chromatographic separation to quantify a specific compound in a complex (often biological) matrix. The three degrees of separation (retention time, precursor *m/z*, and fragment *m/z*) in LC-MRM greatly enhance the S/N enabling targeted quantitation of low-abundance peptides otherwise not detectable in an ambiguous product-ion spectrum (Ekman *et al.*, 2009).

Mass analyzers can also be combined into "hybrid" instruments that unite the advantageous properties of each analyzer they are composed of. For example, a hybrid quadrupole-time-of-flight (QqTOF) instrument combines the scanning speed of a quadrupole with the resolving power of a TOF analyzer to enable high-speed acquisition of spectra at higher resolution. An LTQ-Orbitrap is a hybrid mass spectrometer comprising linear ion trap and Orbitrap analyzers for enhanced resolution and more flexibility. QqTOF and LTQ-Orbitrap instruments have become

the two most prominent high-resolution technologies for proteomics applications for their high resolving power, mass accuracy (stability), linearity, sensitivity, and isotopic fidelity (Eichhorn *et al.*, 2012). Figure 1.13 shows the schematic of TripleTOF™ 5600 (AB Sciex), the hybrid QqTOF used to accomplish the research described in most of this dissertation. Other hybrid MS instruments include quadrupole-ion trap, ion trap-TOF, TOF-TOF, and LTQ-FT-ICR. The latter provides very high resolving power and is especially desired for in-depth structural analysis of intact proteins in top-down proteomics. For example, Lourette *et al.* (2010) used a LC-Q-FT-ICR mass spectrometer to study nitration and oxidation of intact calmodulin as resulted by endogenous oxidative stress.

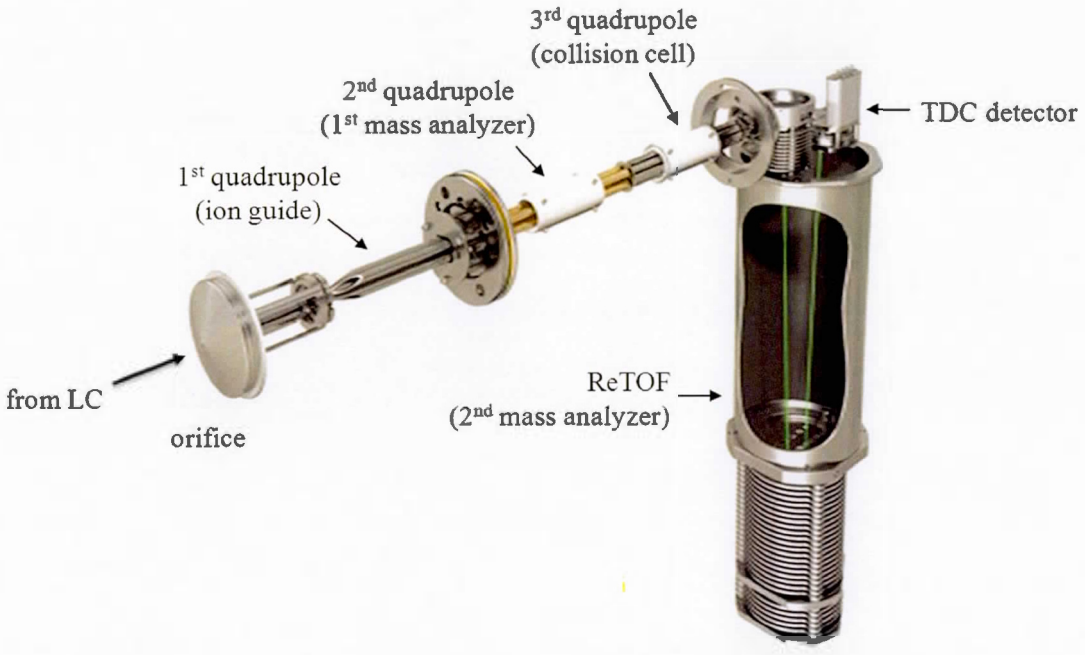


Figure 1.13 General scheme of AB Sciex TripleTOF™ 5600 mass spectrometer, the hybrid QqTOF instrument used in this dissertation. LC: liquid chromatography; ReTOF: reflectron time-of-flight; TDC: time-to-digital converter.

The analysis of complex mixtures requires powerful separation techniques coupled to MS. GC was the first to be coupled with MS and soon became a routine technique for the analysis of mixtures. As discussed in Section 1.2.1, it also became possible in the early 2000s to excise SDS-PAGE or 2D-GE gel spots and subject them to MALDI-MS. However, the application of GC-MS is limited to volatile compounds, and GE-MS was not always feasible. The development of API methods eventually gave rise to more flexible "hyphenated" techniques such as LC-MS, CE-MS[8] and SFC-MS[9] (Gross, 2004). Since most proteins, peptides and metabolites are relatively polar and water-soluble; LC-MS quickly became one of the most efficient and routinely used approaches in proteomics analyses. However, despite the rapid and massive advances in LC technology, current analytical workflows are still not sufficient to fully resolve complex biological samples.

The LC approaches employed in proteomics often aim at separating peptides, for MS-based sequencing, via reversed-phase (RP) chromatography. RP-LC is a relatively fast, effective and reproducible technique; however, most proteomics samples, such as cell lysate digests, are far too complex to be efficiently resolved in a single analytical dimension. The search for a solution to this problem eventually fostered the development of orthogonal multidimensional separation procedures (Di Palma *et al.*, 2012).

1.2.3 Multidimensional Separation in MS-based Proteomics

One of the most powerful two-dimensional (2D) separation techniques was invented in 1975 by O'Farrell and Klose with excellent resolving power for intact proteins, consisting of 2D-GE, and dominated the protein separation field for a long

[8] CE = Capillary Electrophoresis
[9] SFC = Supercritical Fluid Chromatography

time. Another widely used technique, LC, soon gained popularity as it overcame numerous issues associated with GE, including difficulty of automation, lack of reliable quantitation and reproducibility, and especially problematic detection of proteins with large molecular weight, high pI, strong hydrophobicity, or low abundance, such as membrane proteins (Donato *et al.*, 2011).

In 2001, Washburn *et al.* introduced an efficient method for large-scale analysis of yeast proteome using Multidimensional Protein Identification Technology (MudPIT), a 2D-LC-MS/MS approach combining strong cation exchange (SCX) with RP-LC. The goal in multidimensional LC (MD-LC) is to increase peak capacity and thus resolving power to decrease peptide mixture complexity prior to MS/MS analysis. This greatly minimizes "ion suppression" effects, as it dramatically reduces the number of ions entering the mass analyzer at a given time. Ion suppression occurs where highly abundant peptides obscure the detection of other lower-abundance peptides present in the mixture resulting in the loss of information particularly related to less-abundant or less-soluble proteins (Donato *et al.*, 2011). This is likely to be a major challenge in membrane proteomics, where hydrophobic proteins tightly bound within or associated to cell plasma, or organellar, membranes are less amenable to digestion and LC-MS/MS analysis.

Each separation method works based on one or more physicochemical properties. Amino acids have differences in their side chains and thus differ in size, polarity and electrostatic charge. These features can be used as the basis for separation. Peptide separation can therefore take advantage of differences in size (size-exclusion chromatography), polarity (RP and HILIC [10]) or net electrostatic charge (ion exchange and IEF[11]). Some separation techniques combine more than one property to increase selectivity, as "mixed-mode" separations employ different

---

[10] Hydrophilic Interaction Liquid Chromatography
[11] Isoelectric Focusing

analyte-sorbent interaction mechanisms. Electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) (Alpert, 2008), zwitterionic HILIC (Di Palma *et al.*, 2011), and mixed-mode ion-exchange solid-phase extraction (SPE) (Lavén *et al.*, 2009) are examples where peptides are separated based on a combination of more than one physicochemical property. The selectivity of some common techniques that separate peptides based on polarity and charge is demonstrated in Figure 1.14 (Di Palma *et al.*, 2012).

The efficiency of a LC separation is best described by its "peak capacity" (P), theoretically defined as the maximum number of components that could be separated on a LC column within a given time. Peak capacity of a gradient LC separation can be improved by (1) decreasing the gradient slope, (2) increasing the column length with proportional increase in gradient time, and (3) decreasing sorbent particle size (Gilar *et al.*, 2005). Proper use of these strategies has been shown to increase the peak capacity of a single-dimensional (1D) RP-LC from several hundred to a maximum of 1400-1600 (Gilar *et al.*, 2004). Adding another LC dimension to RP-LC can improve the separation power of chromatography tremendously, as the peak capacity of a 2D approach is the linear combination of peak capacities in both separation dimensions:

$$P_{2D} = P_1 \times P_2$$

However, to achieve the maximum peak capacity, the selectivity of the two LC modes must be completely "orthogonal" (complementary). As shown in Figure 1.14, anion/cation exchange chromatography is highly orthogonal with RP-LC and HILIC, and hence often form ideal 2D combinations (Di Palma *et al.*, 2012). SCX-RP-LC is currently the most prevalent MD-LC technique for separating peptides in bottom-up proteomics.

MD-LC can be conducted in "online" or "offline" modes. Online MD-LC, such as MudPIT, involves sophisticated instrumentation and suffers from limited

robustness and flexibility; however, it allows automation, minimal sample loss and contamination, and high-throughput operation. Offline MD-LC is more flexible as it permits the use of incompatible mobile phases in the various LC dimensions. Moreover, offline MD-LC can be performed with larger amounts of sample, and thus is more efficient for the identification of low-abundant proteins, such as membrane proteins (Kong *et al.*, 2011). The SCX-RP-LC approach is compatible with both offline and online configurations. Other separation techniques such as GE, SPE, and affinity-based methods have also been incorporated into MD-LC (Nice *et al.*, 2007).

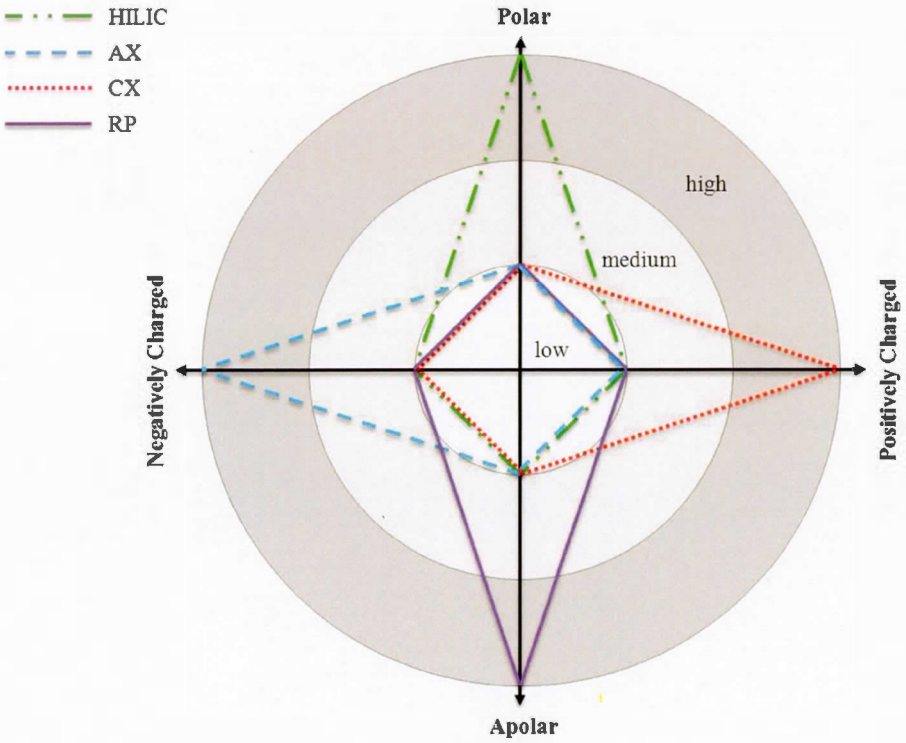


Figure 1.14 Selectivity of HILIC, cation exchange (CX), anion exchange (AX) and RP chromatography for peptides as a function of their charge and polarity indicated as low, medium, high (image adapted from Di Palma *et al.*, 2012)

SPE is one of the most popular sample preparation and enrichment methods, where sample constituents are dissolved in a liquid (mobile phase) and pass through a solid sorbent (stationary phase) followed by elution in a suitable solvent. Similar to chromatography, separation occurs based on the affinity of solutes for the sorbent. SPE disposable cartridges were first introduced in 1978, and have been widely used in various applications ever since. SPE can be performed offline, separated from the subsequent chromatography, or online by direct connection to the GC or LC system (Hennion, 1999). SPE sorbents can be divided in three main groups: (1) inorganic oxides, (2) non-specific sorbents, and (3) compound-specific and class-specific sorbents (Poole, 2003).

Inorganic oxide sorbents mostly include silica, alumina and Florisil (magnesium silicate) and have a high affinity towards water and very polar substances. Non-specific sorbents have a broader range of applications and include surface-modified silica and porous polymers, such as polystyrene-divinylbenzene (PS-DVB) resins and carbon-based materials. Surface-modified silica is the most widely used class of SPE sorbents, although it suffers from several drawbacks including limited stability when subjected to low- or high-pH samples. Nevertheless, these materials are currently the sorbents of choice for extracting analytes with low or moderate polarity from food, environmental and industrial aqueous samples. For more complex matrices, such as biological samples that contain highly polar species or macromolecules, most conventional non-selective sorbents are not appropriate and the use of compound-specific and class-specific sorbents is preferred (Augusto *et al.*, 2013). Compound-specific and class-specific SPE sorbents generally include: (1) surfactant-modified sorbents, (2) mixed-mode polymeric sorbents, (3) molecular recognition sorbents, and (4) nanostructured materials (Augusto *et al.*, 2013). In this dissertation, different mixed-mode polymeric compound-specific/class-specific solid phases were employed using Oasis® (Waters) disposable cartridges. The chemical structure of hydrophilic-lipophilic balanced (HLB), mixed-mode anion exchange

(MAX), and mixed-mode cation exchange (MCX) SPE sorbents and their active sites are depicted in Figure 1.15. As shown, simultaneous presence of aryl and pyrrolidone ring systems in the structure of these sorbents facilitates the retention of amphiphilic compounds such as peptides. However, incorporation of quaternary amine and sulfonic acid active sites in MAX and MCX materials particularly allows for selective retention of peptides with more acidic (rich in Asp and Glu) or basic (rich in Lys, His and Arg) properties, respectively.

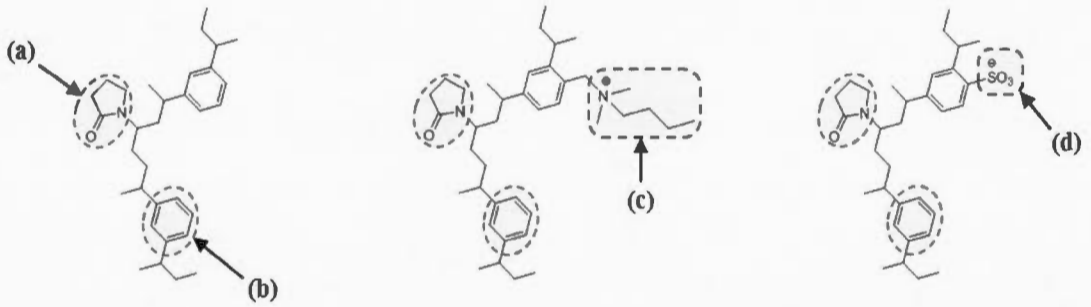


Figure 1.15 Chemical structure of mixed-mode compound-specific/class-specific sorbents used in Oasis HLB, MAX, MCX SPE cartridges showing their hydrophilic interaction (a), hydrophobic interaction (b), anion exchange (c), and cation exchange (d) sites (image adapted from www.waters.com). Polymeric core structure has not been disclosed.

Multidimensional techniques have provided higher peptide separation efficiency and deeper proteome coverage drawing substantial attention during recent years. However, increased time and cost of analysis, lack of complete orthogonality between consecutive dimensions, and more importantly, loss of less-abundant peptides throughout multiple separation steps remain major drawbacks of these methods. These issues and limitations will be discussed in more details in Chapters 2 and 3.

1.3 Challenges and Research Design

The role of reactive metabolite formation and binding to the proteins, and the importance of identifying these adduction targets have been discussed. A brief overview was then given on the experimental techniques often employed for the analysis of proteins and their post-translational modifications in complex mixtures. Several attempts have been made in the past to identify proteins targeted by reactive metabolites (see reviews by Cohen *et al.*, 1997; Zhou, 2003; Zhou *et al.*, 2005; Hanzlik *et al.*, 2009). However, these studies have largely failed to address major challenges in detection of adducted proteins and/or their sites of modification. The difficulties to overcome in these types of studies generally arise from (1) low abundance of modified proteins compared to the large excess of unmodified counterparts, (2) loss of the attached metabolite through hydrolysis or other chemical cleavages caused by the resulting oxidative stress, and (3) low concentration of target proteins in the analyzed sample due to solubility issues, especially in the case of membrane protein targets (Shin *et al.*, 2007; Tzouros and Pähler, 2009).

1.3.1 Preparation of Liver Microsomal Samples

Sample preparation is a key step in proteomics analysis. Denaturing and separating proteins from interfering entities, such as lipids and nucleic acids, their solubilization and efficient digestion, and reducing sample complexity prior to LC-MS/MS analysis significantly enhances data quality. One of the most important groups of proteins that pose technical difficulties at almost each step of the sample preparation procedure is "membrane proteins". These proteins are either physically integrated in a membrane (integral membrane proteins, IMPs) or are chemically linked to the phospholipid structure of the membrane (peripheral membrane proteins,

PMPs). Most proteins need to be active in the cytosol or secreted into the extracellular matrix, and hence are hydrophilic. However, in order to remain stable inside the membrane, IMPs and membrane-associated domains of PMPs are relatively hydrophobic, and have strong intramolecular interactions with membrane lipids through electrostatic forces between the lipid backbone and side chains of hydrophobic amino acid residues within the protein's "transmembrane domains" (TMDs) (Lee, 2003).

The presence of lipophilic residues in TMDs not only greatly reduces membrane protein solubility in aqueous buffers used during sample preparation, but also makes these parts of the protein less accessible to certain proteolytic digestions including widely used trypsin, and endoproteinases Glu-C, Arg-C and Lys-C, which cleave polypeptide chains at hydrophilic residues. Table 1.7 lists the site-specificity of proteases most commonly used in bottom-up proteomics (Westermeier and Naven, 2002).

Table 1.7 Proteases commonly used in bottom-up proteomics

Protease	Cleavage Site [1]	Exception
Trypsin	Arg, Lys	Adjacent to Pro
Endoproteinase Glu-C	Asp, Glu	Adjacent to Pro
Endoproteinase Arg-C	Arg	Adjacent to Pro
Endoproteinase Lys-C	Lys	Adjacent to Pro
Chymotrypsin[2]	Ile, Leu, Met, Phe, Trp, Tyr, Val	Adjacent to Pro
Elastase[2]	Ala, Gly, Ile, Leu, Ser, Val	
Pepsin[2]	Glu, Leu, Phe	

[1] Cleavage occurs at the *C*-terminal side of the given residues

[2] Enzymes with low cleavage site-specificity

Trypsin is a selective endopeptidase that cleaves at the *C*-terminal side of Arg and Lys residues. However, Burkhart *et al.* (2012) demonstrated that trypsin's specificity, efficiency, reproducibility and its potential impact on quantitation and proteome coverage can considerably vary depending on the source and grade of the enzyme. For large-scale studies, it is often preferable to use a less expensive commercial version of trypsin, even with the caveat of less specificity, since otherwise the cost of the study would not be feasible for most research groups. Figure 1.16 illustrates the amino acid sequence of a typical IMP (Scheuring *et al.*, 1999).

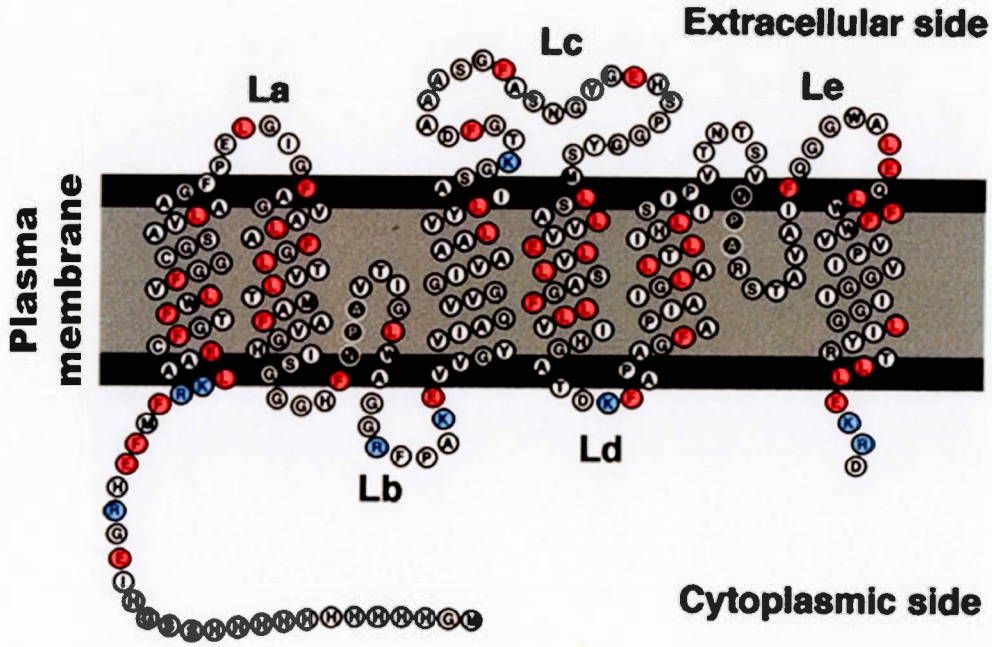


Figure 1.16 Amino acid sequence and approximate topology of *Escherichia coli* water channel aquaporin Z and its transmembrane regions. Trypsin enzymatic cleavage sites are coloured in blue. The residues in red mark potential pepsin cleavage sites which targets more hydrophobic residues (image adapted from Scheuring *et al.*, 1999).

As depicted in Figure 1.16, trypsin cleavage sites are usually located on the cytosolic and extracellular regions of an IMP. Therefore, even if protein-lipid interactions are broken and TMDs are fully solubilized, a generic trypsin digestion procedure would likely not provide enough sequence coverage from the transmembrane regions. However, if a protease with more specificity towards hydrophobic residues is employed in conjunction; better sequence coverage could be obtained for such a protein. This idea led to the development of a dual-enzyme digestion approach for more efficient analysis of liver microsomes, which are rich in membrane proteins, in the first part (Chapter 2) of this dissertation.

For protein solubilization, two types of intramolecular interactions must be overcome: (1) protein-protein interactions, and (2) protein-lipid interactions. Ionic detergents such as alkyl sulfates, alkyl phosphocholines, and bile acid salts can efficiently disrupt protein-protein interactions and are commonly used in most proteomics applications. Non-ionic detergents, on the contrary, readily interfere with protein-lipid and lipid-lipid interactions and are more preferred for samples enriched in membrane proteins, such as the microsomal fractions. Chaotropes, such as urea, thiourea and guanidinium chloride, are strong denaturing agents that stabilize unfolded protein states via hydrogen bonds and electrostatic interactions and are effective against all above-mentioned interactions. Some organic solvents and carboxylic acids have intrinsic protein denaturing properties and disruptive effect on membrane bilayers thus enhancing IMP solubilization (Speers and Wu, 2007). In this thesis, members from each group of solubilizing agents were tested for the analysis of liver microsomal proteins, and the method with the highest performance was selected (Figure 1.17). Optimization of the solubilization and digestion steps for the analysis of rat liver microsomes is discussed in more detail in Chapter 2.

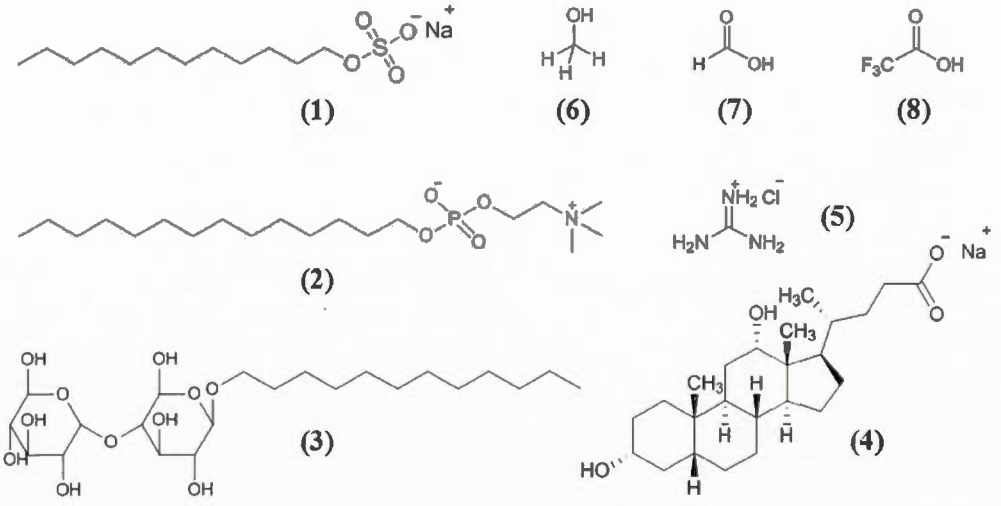


Figure 1.17 Solubilizing agents tested for the analysis of liver microsomal proteins were selected from different classes: (1) sodium *n*-dodecyl sulfate (SDS); (2) *n*-tetradecyl phosphocholine (FC-14); (3) *n*-dodecyl maltoside (DDM); (4) sodium deoxycholate (SDC); (5) guanidinium hydrochloride; (6) methanol; (7) formic acid; (8) trifluoroacetic acid.

Liver microsomes are often used to study xenobiotic metabolism *in vitro*, and covalent binding to microsomal proteins serves as a surrogate marker for toxicity mediated by reactive metabolites (Shin *et al.*, 2007). Microsomes are membrane vesicles formed in subcellular fractionation, after separation of S9 and cytosolic fractions, consisting largely of the endoplasmic reticulum (ER), plasma membranes, lysosomes, peroxisomes, nuclear membranes, and cytoplasm, and are rich in CYP, UGT, GST, FMO,[12] and other xenobiotic-metabolizing enzymes. *In vitro* microsomal metabolism studies provide useful information on oxidative pathways involved in reactive metabolite formation and covalent binding to proteins, which are, to some extent, comparable to the tendency for covalent binding and toxicity *in vivo* (Shin *et*

[12] Flavin-Containing Monooxygenase

al., 2007). The workflow used to identify potential target proteins of reactive metabolites from liver microsomal incubations is demonstrated in Figure 1.18.

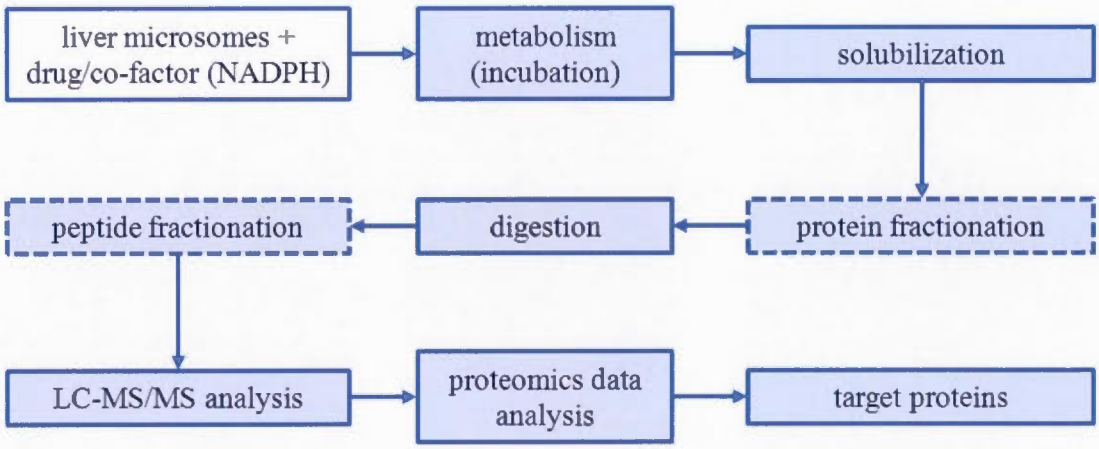


Figure 1.18 Workflow of an *in vitro* experiment using liver microsomes to study protein covalent binding

Microsomal samples contain large quantities of impurities and compounds that can hinder protein digestion and/or downstream LC-MS/MS analysis, and hence require post-incubation clean-up procedures. Interfering compounds can decrease the efficiency of RP-LC separation via selectively binding to the active sites of the stationary phase ("column saturation") and also cause ion suppression, affecting the MS performance (Polettini, 2006). The main cause of ion suppression in ESI-MS of the biological samples is a change in the spray droplet solution properties caused by the presence of non-volatile or less volatile solutes including salts, ion-pairing agents, endogenous compounds, drugs and metabolites. These compounds change the efficiency of droplet formation or droplet evaporation affecting the amount of ions that eventually reach the detector. It has also been shown that: (1) analytes with similar or close *m/z* values can suppress each other; (2) molecules with higher mass

can suppress the signal of smaller molecules; and (3) more polar analytes are more susceptible to ion suppression (Annesley, 2003). To minimize potential LC-MS interferences from microsomal samples, several clean-up methods were tested in this work including protein precipitation, molecular weight cut-off filtration and mixed-mode SPE. These efforts will be described in details in Chapters 2, 3 and 5.

1.3.2 Data Processing Workflow

Another challenge in modern MS-based protein identification is effectively using database search software to figure out the correct peptide spectral match while maintaining control over false identifications ("false positives"). Most protein search engines provide an estimated probability of correctness of putative identifications through "false discovery rate (FDR)" analysis. These analyses usually employ a technique known as "target-decoy database searching", in which the database search engine is provided with potential incorrect answers (decoys) generated by randomizing or reversing true sequences (targets). The rate of occurrence of these known-incorrect answers in the search output is then used to estimate the FDR (Tang *et al.*, 2008). False protein IDs usually originate from low-confidence peptide spectral matches (inaccurate precursor mass or low-quality MS/MS spectra), or wrong modification identification (Shin *et al.*, 2007).

Although state-of-the-art software produce lists of identified proteins at as low as 1% FDR, false positives can never be fully eliminated and often require manual verification of the corresponding LC-MS/MS data. Other efficient strategies to efficiently reduce FDR are (1) increasing the number of analytical replicates (n), and (2) the use of control samples (Veenstra and Smith, 2003). In this thesis, LC-MS/MS data from acetaminophen-treated rat liver microsomes were compared to control (no drug) samples using multiple complementary data analysis approaches. The final list

of putative hits was then manually inspected in terms of accurate mass, charge state, and MS/MS spectra to exclude remaining potential false positives. The developed data processing workflow will be widely discussed in Chapter 5.

1.3.3 Research Outline

This research was originally designed in three phases attempting to address important challenges in the proteomic workflow for identifying reactive metabolite protein targets. The main focus of Phase I was on protein solubilization and digestion. In this phase, several solubilizing agents including detergents, chaotropes, methanol and acids were tested for their capability to extract proteins from liver microsomal samples. In addition, different proteases were examined separately, and in combination, under basic and acidic conditions. The optimal solubilization-digestion scenario was selected based on performance characteristics, that is, the number of identified peptides and proteins, and average protein sequence coverage. An initial offline 2D-LC-MS/MS approach was also developed coupling SCX peptide fractionation to ultra-high-pressure RP-LC prior to the MS/MS analysis. The outcome of this part of the study was published as a research article and is presented in Chapter 2.

In phase II, the objective was to maximize the separation power of the initial LC-MS/MS method for the efficient analysis of protein-enriched liver samples. Protein- and peptide-level ion-exchange fractionation methods were compared, separately and combined, to regular (no fractionation) LC-MS/MS and the method with the best performance was selected and tested for liver S9 and microsomal fractions from rat, mouse, and human. The method development and cross-species analysis results were published in two articles and will be discussed in Chapters 3 and 4.

Finally, lessons learned from Phases I and II were applied to real biological samples in attempt to identify protein targets of acetaminophen's reactive metabolite following *in vitro* microsomal incubations. For an in-depth data processing of the results, a four-stage data analysis workflow was established for the identification of target proteins while focusing on removing false positives (Chapter 5). In the last chapter, a semi-targeted approach is presented for detecting covalent binding of a specific protein (MMP-13) *in vitro* and validated using patient chondrocyte cell cultures. Figure 1.19 demonstrates the general layout of this dissertation and challenges addressed in each phase.

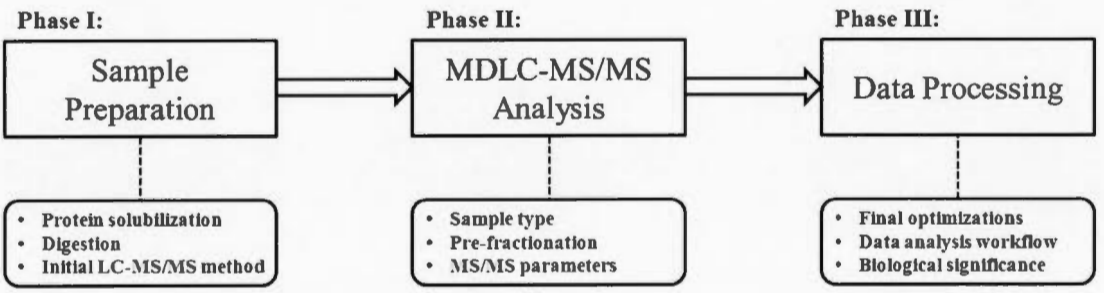


Figure 1.19 Research design in the development of analytical methods for the identification of reactive metabolites target proteins

CHAPTER TWO

OPTIMIZED PROTEOMIC ANALYSIS OF RAT LIVER MICROSOMES USING DUAL ENZYME DIGESTION WITH 2D-LC-MS/MS

Makan Golizeh and Lekha Sleno

Published in **Journal of Proteomics 2013, 82, 166-178**. Supplementary data available online at doi:10.1016/j.jprot.2013.02.001.

The first step towards the identification of reactive metabolite protein targets was to develop an analytical method to maximize the number and sequence coverage of the proteins identified in liver microsomes. In this chapter, an optimized sample preparation workflow was established with a focus on solubilization and digestion of rat liver microsomal proteins using 2D-LC-MS/MS. SDS-assisted solubilization with parallel trypsin/pepsin dual-digestion combined with SPE clean-up, SCX chromatography and RP-UHPLC-HR-MS/MS provided an efficient approach for the analysis of rat liver microsomes. The sample preparation method developed in this chapter was improved and used to analyze liver proteins in rat, mouse and human in Chapters 3 and 4, as well as the acetaminophen-treated samples in Chapter 5.

Makan Golizeh and Lekha Sleno are co-authors of this article. Makan Golizeh conducted the literature survey, prepared experimental protocols, conducted the experiments, processed the data, and prepared the original manuscript. Professor Lekha Sleno supported and supervised the project, verified data analysis and interpretation of the results, revised and finalized the manuscript.

2.1 Abstract

A systematic approach was developed to optimize the analysis of rat liver microsomes combining ion exchange fractionation with reversed-phase chromatography coupled to high resolution quadrupole-time-of-flight mass spectrometry. A comparison was performed with several conditions to select the most efficient solubilization and proteolysis protocol to achieve highest proteome coverage. Optimal trypsin digestion conditions were achieved with SDS and heat to increase solubilization of microsomal samples, with an increase from 621 to 686 identified proteins when SDS and heat were applied. Pepsin digestion yielded complementary results, especially in terms of hydrophobic environments, thus allowing sequence coverage to be increased substantially. Several dual digestion strategies were tested, with trypsin and pepsin combined in series or in parallel. A parallel tryptic-peptic dual digestion, combining mass spectral data of single enzyme digestions, yielded the best results in terms of number of identified proteins, increasing by 29% from the best single enzyme procedure, and sequence coverage improved by 5% on average for all proteins identified. Using our complete set of data, a total of 1095 proteins were identified with less than 1% FDR, out of which 213 proteins (19.5%) were integral membrane proteins. Proteomics data have been deposited to the ProteomeXchange Consortium with dataset identifier **PXD000128**.

2.2 Introduction

Approximately 25% of all translated genes code for membrane proteins, possessing two or more transmembrane domains, and this value increases to 40%, if lipid-anchored or membrane-associated proteins and those having a single transmembrane helix are taken into account (Bendz *et al.*, 2009). Membrane proteins have strong implications in cell survival, with significant roles in cellular events such

as molecule transport, signal transduction, cell–cell interactions, and response to changes in environmental conditions (Rietschel, Bornemann, *et al.*, 2009; Vertommen *et al.*, 2011). They are an extremely valuable class of proteins since they represent the targets of 70–80% of all drugs (Mathias *et al.*, 2011), among which G-protein coupled receptors (GPCRs), protein kinase, cytochrome P450, syntaxin, prostaglandin G/H synthase 1 (PTGS) or cyclooxygenase (COX) families are some remarkable examples (Imming *et al.*, 2006). Two main categories of membrane proteins exist; those which are physically integrated into the membrane, tightly bound and only released using strong solubilizing agents (integral membrane proteins, IMP) and those that are associated with the membrane and more easily released from the membrane using gentle extraction procedures (peripheral membrane proteins, PMP) (Vertommen *et al.*, 2011). Integral membrane proteins (IMPs) can be either β-barrels or α-helices. α-Helix bundle proteins are the most abundant membrane proteins, having one or more α-helical transmembrane domains (TMDs), and pose technical difficulties for proteomic characterization due to their hydrophobic nature. These proteins can be either monotopic (membrane-anchored), bitopic (single-pass integral), or polytopic (multiple-pass integral) (Gilmore and Washburn, 2010; Speers and Wu, 2007).

Because of their solubility issues, membrane proteins can be absent from a simple cell lysate and thus protein enrichment is essential for their analysis. Common strategies include: subcellular fractionation, delipidation and affinity purification (Gilmore and Washburn, 2010). In subcellular fractionation, which is most often employed, low abundant hydrophobic proteins can be enriched into a microsomal fraction (Van Summeren *et al.*, 2012). Microsomes are small lipid vesicles which form spontaneously when cells are mechanically disrupted, and are a rich source of IMPs from endoplasmic reticulum (ER), plasma membrane, mitochondria, and Golgi apparatus (Mathias *et al.*, 2011). Liver microsomes are of particular interest, since they contain many metabolic enzymes for drugs and endogenous compounds and are

extensively used in pharmacokinetic and drug metabolism studies (Huang *et al.*, 2011). Numerous studies have been performed on rat (Huang *et al.*, 2011; Galeva and Altermann, 2002; Galeva *et al.*, 2003; Arnold *et al.*, 2004; Nisar *et al.*, 2004), mouse (Mathias *et al.*, 2011; Kanaeva *et al.*, 2005; Zgoda *et al.*, 2009; Peng *et al.*, 2012) and human (Seibert *et al.*, 2009; Petushkova *et al.*, 2006; Yukinaga *et al.*, 2012) liver microsomes to study proteins involved in xenobiotic metabolism and drug-induced hepatotoxicity.

In addition to subcellular fractionation, recent improvements in hybrid separation techniques have led to the expansion of shotgun proteomics and helped overcome some of the challenges in studying membrane proteins. Two-dimensional techniques using complementary separations can be employed to increase the coverage of the proteome (Gilmore and Washburn, 2010). The most common strategy remains the combination of strong cation exchange (SCX) (on-line or off-line) with reversed-phase (RP) separation. An automated online SCX-RP-based methodology, known as Multidimensional Protein Identification Technology (MudPIT), has been well established for the analysis of comprehensive protein expression profiles (Washburn *et al.*, 2001). However, in comparison with online 2D-LC-MS, which is faster and less laborious for high-throughput proteomic studies, an offline approach provides increased flexibility and higher resolution due to the continuous linear gradient of SCX-chromatography in the first dimension without intermittent RP-chromatography separations (Nägele *et al.*, 2004). The primary role of any pre-fractionation is to reduce the complexity of the peptide mixture, so that the mass spectrometer has sufficient time to fully interrogate the sample, with the added benefit of decreasing ion suppression (Khan *et al.*, 2011).

Mass spectrometry-based proteomics platforms include quadrupole-time of flight (QqTOF), linear ion trap, Orbitrap, and Fourier transform ion cyclotron resonance (FT-ICR) instruments (Walther and Mann, 2010). Orbitraps and QqTOF

instruments in combination with liquid chromatography have become the two most prominent systems used in proteomics, with high resolution capabilities for accurate mass measurements as well as high sensitivity and isotopic fidelity (Eichhorn *et al.*, 2012). The most recently introduced platforms have also increased their sequencing speeds so many more eluting peaks can be targeted for MS/MS in a data-dependent fashion, thus increasing the potential for higher proteome coverage.

To achieve good sequence coverage, proteins must be fully accessible for digestion, which remains a significant challenge for the analysis of membrane proteins (Griffin and Schnitzer, 2011). Solubilizing agents thus facilitate sample preparation in membrane proteomics, and can be divided into several categories (Speers and Wu, 2007; Berridge *et al.*, 2011; Chen *et al.*, 2007; Le Maire *et al.*, 2000). Chaotropes are strong denaturing agents that stabilize unfolded proteins via hydrogen bonds and electrostatic interactions. Linear-chain ionic detergents (e.g. SDS) are extremely efficient at solubilizing proteins. Bile acid salts are steroidal ionic detergents and are significantly milder solubilizing agents. Non-ionic detergents have polyoxyethylene (POE) or glycosidic polar groups combined with a hydrophobic chain and are also considered to be mild detergents. Zwitterionic detergents also exist, including sulfobetaine and phosphocholine derivatives. MS-compatible detergents have been developed to solve the interference issue of many detergents with mass spectrometry. These detergents either easily break down into non-surfactant by-products or do not interfere with proteomic analysis due to their orthogonal LC elution (high organic %) to most peptides. Aqueous-organic solvent systems are another alternative for facilitating protein digestion and are directly compatible with LC-MS. Finally, organic acids can also be useful for membrane disruption and integral membrane protein solubilization. SDS has been the most commonly used detergent for the analysis of liver microsomes (Huang *et al.*, 2011; Galeva and Altermann, 2002; Galeva *et al.*, 2003; Nisar *et al.*, 2004; Zgoda *et al.*, 2009; Peng *et al.*, 2012; Seibert *et al.*, 2009; Petushkova *et al.*, 2006). However,

Triton X-114 (Mathias *et al.*, 2011), sodium cholate (Kanaeva *et al.*, 2005) and the MS-compatible RapiGest (Arnold *et al.*, 2004) have also been employed.

Even though solubilized membrane proteins provide better access to trypsin, TMDs often lack tryptic cleavage sites; therefore the use of this enzyme has been challenged for membrane proteins. Consequently, less specific proteases like chymotrypsin (Speers and Wu, 2007; Yu *et al.*, 2003), elastase (Rietschel, Arrey, *et al.*, 2009) and proteinase K (Bendz *et al.*, 2009; Speers and Wu, 2007) have been considered as alternatives. Aside from these, pepsin has also been suggested as a valuable tool in membrane proteomics. It is presumed that, due to its more hydrophobic cleavage sites, it can better cleave membrane-embedded lipophilic portions of IMPs leading to a better sequence coverage for the analysis of these proteins (Rietschel, Bornemann, *et al.*, 2009; Han and Schey, 2004). However, trypsin offers some major advantages for LC-MS/MS analysis, not easily ignored even in the case of membrane proteins, such as an optimal average peptide length of ~ 14 amino acids and the presence of at least two positive charges, at the *N*-terminus and *C*-terminal Arg or Lys (Burkhart *et al.*, 2012). Dual enzyme digestion procedures have been employed to combine the benefits of complementary cleavage behaviors, increasing protein sequence coverage by generating more analyzable peptides per protein. Trypsin-chymotrypsin (Fischer and Poetsch, 2006), trypsin-endoproteinase Glu-C (Prabakaran *et al.*, 2001), trypsin-proteinase K (Distler *et al.*, 2006; Han *et al.*, 2002), trypsin-cyanogen bromide (Washburn *et al.*, 2001; Fischer and Poetsch, 2006; Van Montfort *et al.*, 2002), proteinase K-cyanogen bromide (Speers *et al.*, 2007), and pepsin-pancreatin (Reyes *et al.*, 2006) are some examples of sequential dual enzyme digestions. It is also been suggested in a previous article (Mbeunkui and Goshe, 2011), where they tested four different and complementary solubilization techniques for microsome analysis, that applying two different solubilization methods would provide a more effective coverage of membrane proteins. Therefore, if using a dual enzyme system for increasing coverage of the rat microsome proteome, different

solubilizing agents added at each digestion step could potentially ameliorate the results even further.

In this study, offline 2D-LC was coupled to a high-resolution QqTOF mass spectrometer for the analysis of rat liver microsomes with combined tryptic and peptic digestions, using SDS and MeOH/TFA as complementary solubilizing methods. Results were compared to those using solubilizing agents from different classes with single enzymes, as well as sequential and parallel dual digestions in terms of number of peptides and proteins, integral membrane proteins, and overall sequence coverage.

2.3 Experimental

2.3.1 Materials

Male Sprague-Dawley rat liver microsomes (20 mg/ml of protein) were purchased from Celsis (Baltimore, MD). Sequencing-grade modified trypsin was obtained from Promega (Madison, WI). Iodoacetamide and trifluoroacetic acid were from Alfa Aesar (Ward Hill, MA), trichloroacetic acid from VWR International (West Chester, PA), and *n*-tetradecyl phosphocholine (FC-14) from Affymetrix (Maumee, OH). Porcine gastric mocusa pepsin and type XVII-B endoproteinase Glu-C from *Staphylococcus aureus* strain V8, as well as all other chemical reagents (analytical grade) were purchased from Sigma-Aldrich (St. Louis, MO). HPLC-grade methanol and acetonitrile were acquired from EMD Chemicals (Gibbstown, NJ). All solutions were prepared using nanopure water purified by a Millipore Synergy UV system (Molsheim, France).

2.3.2 Sample Preparation

Rat liver microsomes (500 µg, 25 µl) were added to a solubilizing solution (40 µl) and sonicated for 10 minutes (Branson 2510 ultra-sonic bath, Danbury, CT). Samples labeled as **TO** (trypsin only), **GC** (Glu-C) and **P** (pepsin) were suspended in 100 mM ammonium bicarbonate pH 8.5 For other samples, the following solutions were used for solubilization: 4% sodium *n*-dodecyl sulfate (SDS) for **TS** (trypsin with SDS), **TSH** (trypsin with SDS and heat) and **PSH** (pepsin with SDS and heat), 1% sodium deoxycholate (SDC) for **TB** (trypsin with bile acid), 2% *n*-dodecyl maltoside (DDM) for **TD**, 1% *n*-tetradecyl phosphocholine (FC-14) for **TF**, 6 M guanidinium chloride for **TG** and 60% methanol in 20 mM Tris-HCl for **TM**. **TSH** and **PSH** were incubated for 3 minutes at 95°C. All conditions were performed in duplicate. Samples were subsequently incubated with dithiothreitol (DTT) at a final concentration of 25 mM to reduce disulfide bonds (10 min, RT). Reduced samples were then alkylated with iodoacetamide (IAM) at a final concentration of 50 mM for 30 minutes at 37°C in the dark. Protein precipitation was performed with 20% trichloroacetic acid (TCA) and leaving on ice for 15 minutes. After centrifugation (13,000 rpm, 10 min at 4°C), the supernatant was discarded and pellets were washed with 10 % TCA (150 µl), centrifuged (13,000 rpm, 5 min), then washed with nanopure water (100 µl) three times to remove residual acid.

2.3.3 Digestion Conditions

Protein pellets were resuspended in 100 mM ammonium bicarbonate, pH 8.5 (250 µl) for trypsin and Glu-C digestions, or in 0.11% trifluoroacetic acid in 10% methanol, pH 2.0 (500 µl) for pepsin digestions, and incubated at a 1:50 (w/w) enzyme to protein ratio for 18 h, at 37°C for trypsin and Glu-C and at 25°C for

pepsin. Digestion was quenched by adding 1% formic acid (50 µl) (for trypsin and Glu-C) or 1% ammonium hydroxide (50 µl) (for pepsin) and protein digests were evaporated to dryness under vacuum (ThermoFisher Scientific Universal Vacuum System, Asheville, NC). For sequential digestions (**T-P**, **P-T**, **T-C**, **TSH-P** and **PSH-T**), the resolubilized sample was subjected to a second treatment with either pepsin for 4 h, trypsin for 18 h, or cyanogen bromide for 15 h. In the case of **T-C**, the tryptic digest was incubated first with 90% formic acid (100 µl) at room temperature for 5 minutes, followed by the addition of 100 mg cyanogen bromide and the reaction stopped by adding methanol (250 µl). Resulting samples were also evaporated to dryness. All samples were kept at -30°C prior to SCX fractionation.

2.3.4 Strong Cation Exchange Fractionation

Protein digests were reconstituted in buffer A (100 µl), sonicated for 10 minutes, and centrifuged (13,000 rpm, 5 min) prior to injection (40 µl) onto a Zorbax 300-SCX 2.1 x 150 mm column with 5 µ particles (Agilent Technologies, Palo Alto, CA) using an Agilent 1200 series HPLC equipped with a binary pump, degasser, diode array detector and fraction collector. Strong cation exchange (SCX) fractionation was performed at a flow rate of 300 µl/min using a gradient of 0-50% B in 15 min, up to 100% B at 25 min, then held for an additional 7 min at 100% B, where buffers A and B were 10 mM potassium dihydrogen phosphate in 25% acetonitrile (pH 2.75), and 1 M potassium chloride in buffer A (pH 2.75), respectively. Absorbance was monitored at 220 and 280 nm. One minute fractions (300 µl) were aliquoted into a 96-well plate. For tryptic digestions, fractions eluting from 5-21 min were subsequently evaporated, whereas for peptic digestions, fractions eluting from 4-23 min were used (Figure 2.1). In the case of parallel dual digestions, tryptic and peptic digests were mixed and then fractionated (**TP-mf**); with fraction

collection performed as for samples from trypsin-only digests. For **TP-fm** samples, SCX fractions were combined from 16 tryptic fractions and 19 peptic fractions, by merging three pairs of peptic fractions with the lowest UV signals, in order to have a total of 16 samples for LC-MS/MS analysis.

2.3.5 Reversed-Phase UHPLC-MS/MS

Dried fractions were reconstituted in 10% acetonitrile (100 µl) and injected (20 µl) onto a Kinetex XB-C_{18} 2.1 x 100 mm column, with solid core 1.7 µm particles (100 Å) (Phenomenex, Torrance, CA) using a Nexera UHPLC (Shimadzu, Columbia, MD). Reversed-phase liquid chromatography was performed using 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B), with a gradient of 5-30% B in 15 min, to 40% B at 24 min, followed by a sharp increase to 80% B at 25 min and held for an additional 3 min prior to re-equilibration of the column, and a flow rate of 300 µl/min at 40°C. To protect the ion source against high salt concentrations from SCX elution, a Rheodyne switch valve (IDEX Health and Science, Oak Harbor, WA) was used to send eluent to waste during the first two minutes of the run. All MS and MS/MS spectra were collected on a high-resolution hybrid quadrupole-time-of-flight (QqTOF) TripleTOF 5600 mass spectrometer (AB Sciex, Concord, ON, Canada) equipped with a DuoSpray ion source in positive ion mode. The instrument performed a survey TOF-MS acquisition from *m/z* 140-1250 with an accumulation time of 250 ms, followed by MS/MS on the ten most intense ions with +1 to +4 charge states from *m/z* 250-1250 using information-dependent acquisition (IDA) with dynamic background subtraction (DBS). Each MS/MS had an accumulation time of 100 ms and collision energy ranging from 20 to 40 V, depending on the charge state and mass of the product ion. The total cycle time was 1.3 seconds. Figure 2.1 depicts representative LC-MS/MS data from tryptic (**TSH**) and peptic (**P**) digestions.

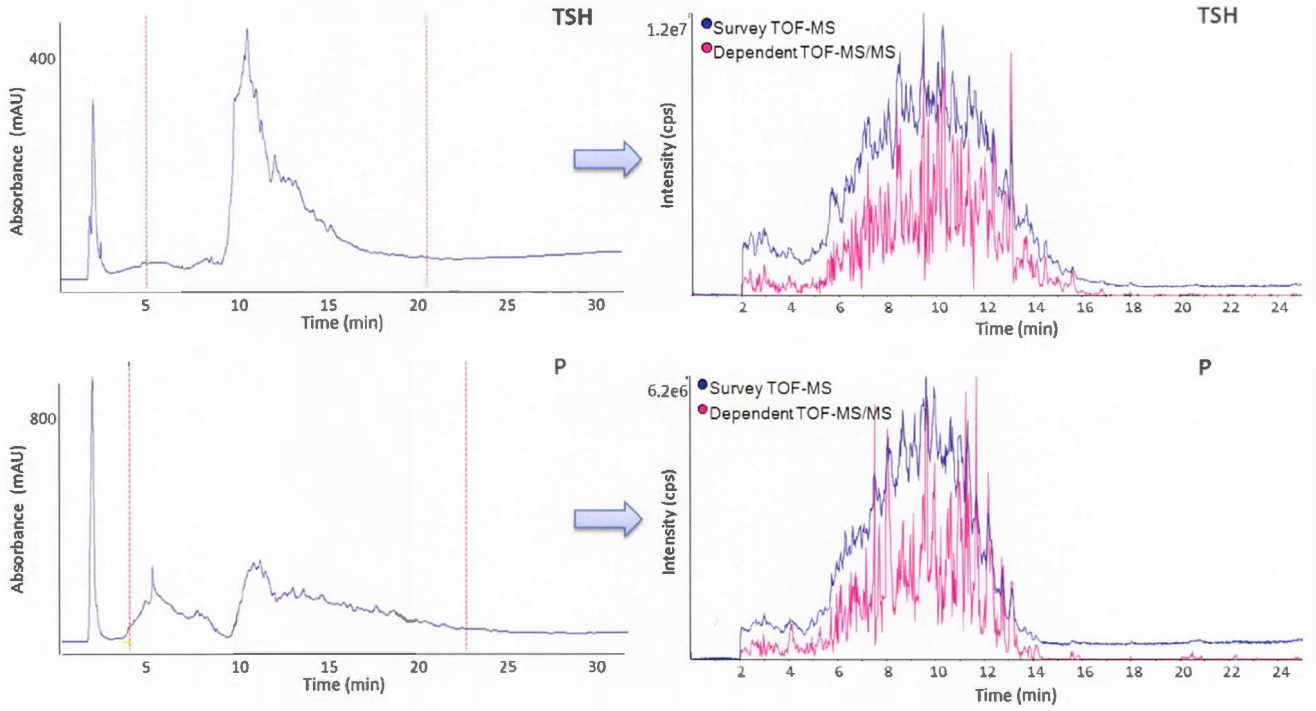


Figure 2.1 Representative UV chromatogram (detected at 220 nm) from SCX fractionation (left) and total ion chromatograms from LC-MS/MS analysis of one SCX fraction showing survey TOF-MS and sum of dependent MS/MS scans (right) from trypsin (**TSH**) and pepsin (**P**) digestions. The portion of SCX elution used for LC-MS/MS analysis is shown with dotted lines.

2.3.6 Data Processing

MS/MS files from each set of fractions were concatenated and searched against the UniProt protein database (release date 12-07-2011, 241 MB) by ProteinPilot software (version 4.1) using Paragon algorithm (Shilov *et al.*, 2007) with the following parameters: protein identification with thorough ID search effort, iodoacetamide for cysteine alkylation and no specified proteolytic enzyme for *rattus norvegicus* with ID focus on biological modifications, detection protein threshold of unused ProtScore > 0.05 (confidence > 10.0%) and including false discovery rate (FDR) analysis. MS tolerance was 0.05 Da on precursor ions and 0.1 Da on fragments and the search was performed for +2 to +4 charge states. For **TP-co**, MS/MS files from **TSH** and **P** were co-processed representing a sample in which SCX fractionation and RP-LC steps have been performed separately and data then combined for searching purposes for two parallel tryptic and peptic digestions. Using ProteinPilot Descriptive Statistics Template (version 3.001p) (http://www.absciex.com/PDST), proteins were identified based on 1% global false discovery rate criteria using a target-decoy database search algorithm (Tang *et al.*, 2008). ProteinPilot Protein Alignment Template (version 2.000p) was also used for replicate analysis. MS spectra for **TSH**, **P**, **TP-co** data sets have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaino *et al.*, 2010) with the dataset identifier **PXD000128**.

The list of UniProt accession numbers was uploaded to NCBI Batch Entrez (http://www.ncbi.nlm.nih.gov/sites/batchentrez) to obtain the batch FASTA file, which was subsequently submitted to ExPASy (http://www.expasy.org) for determination of isoelectric point and monoisotopic molecular weight and to Phobius (http://phobius.sbc.su.se) for the prediction of integral membrane proteins, and compared to predictions done by other traditional transmembrane predictive

algorithms such as TMHMM (http://www.cbs.dtu.dk/services/tmhmm), SCAMPI (http://scampi.cbr.su.se), and SOSUI (http://bp.nuap.nagoya-u.ac.jp/sosui). GRAVY Calculator (http://www.gravy-calculator.de) was used to calculate grand average of hydropathy (GRAVY) scores. The list of accession numbers was also uploaded to PANTHER (Thomas *et al.*, 2003) (http://www.pantherdb.org) for gene ontology classification and to InnateDB (Lynn *et al.*, 2008) (http://www.innatedb.ca) in order to perform the protein-protein interaction analysis with orthologous UniProt identifiers as the cross-reference database and filtered to show only interactions between uploaded proteins. The results were subsequently visualized with the Cerebral plugin (Barsky *et al.*, 2007) of Cytoscape software (http://www.cytoscape.org) version 2.8.2 to organize the interaction network in the context of subcellular localization for all of the proteins displayed and visualized by Cerebral. A Venn diagram was created by BioInfoRx area-proportional Venn diagram plotter (http://apps.bioinforx.com) and other data were processed and visualized by Microsoft Excel 2010.

2.4 Results and Discussion

In this comparative study, eighteen different treatment conditions, including three different enzymes and seven solubilizing agents, were tested in order to select the enzyme-detergent combination with the highest proteome coverage, using single enzyme, sequential dual enzyme, and parallel dual enzyme digestion procedures. A thorough analysis of all proteins identified was then performed to describe the rat liver microsomal proteome based on several physicochemical characteristics, including isoelectric point, molecular weight and hydropathy distributions, as well as the extent of sequence coverage of these proteins and their functional roles.

2.4.1 Single-Enzyme Digestion

Trypsin, being the most commonly used protease for shotgun proteomics application, but with certain limitations for membrane proteins, was tested in the presence of multiple solubilizing agents to increase proteome coverage. Seven different categories of solubilizing agents were used, in order to assess all available classes of molecules. Also, in terms of enzymatic proteolysis, since microsomal fractions are rich in membrane proteins, it was anticipated that a less-specific enzyme, such as pepsin, would help identify more IMPs by cleaving membrane-embedded portions normally inaccessible to trypsin. As expected, under detergent-free conditions (**TO** *vs.* **P**), pepsin digestion identified a higher proportion of IMPs while trypsin led to a higher number of proteins and a better sequence coverage overall. Table 2.1 gives the compiled results of all tested conditions from this study in terms of number of proteins and peptide identified, sequence coverage and proportion of IMPs. Glu-C did not provide satisfactory results under the conditions tested in comparison to the other two enzymes. Figure 2.2 compares the percentage of identified proteins for trypsin, pepsin and endoproteinase Glu-C digestions of the total number of proteins covered by these three enzymes without added solubilizing agent (786 proteins in total). The Venn diagram clearly demonstrates that Glu-C adds the least amount of complementary data, and shows that the combination of pepsin and trypsin covers the vast majority of proteins (95%). Glu-C digestion also yielded a drastically lower number of peptides than the other two enzymes tested, and thus was not chosen for further evaluation.

Table 2.1 Comparison between different digestion conditions for analysis of rat liver microsomes by 2D-LC-MS/MS

Sample [a]	Avg. # Proteins (%CV) [b]	Reproducibility [c]	# Proteins [d]	# Peptides [d]	Avg. Sequence Coverage (%) [d]	# IMP (%IMP) [d, e]
TO	621 (5.7)	76%	693	4489	19.7	131 (19.0)
P	345 (10.8)	72%	402	2111	15.8	78 (19.5)
GC	279 (1.0)	73%	315	834	10.9	48 (15.5)
TS	609 (3.7)	81%	667	4171	20.1	122 (18.4)
TSH	686 (1.3)	81%	750	5278	21.0	140 (18.7)
TB	613 (1.4)	79%	674	4447	21.1	119 (17.8)
TD	594 (9.9)	78%	660	4237	21.5	124 (18.9)
TF	560 (6.3)	77%	625	4081	20.1	118 (19.0)
TG	558 (6.6)	78%	618	3749	19.0	119 (19.3)
TM	601 (1.4)	83%	626	4351	21.6	109 (17.5)
T-P	471 (2.4)	75%	538	2929	13.5	100 (18.7)
TSH-P	368 (0.0)	79%	403	1801	12.5	72 (18.0)
P-T	513 (14.5)	71%	597	4481	20.5	119 (20.0)
PSH-T	531 (3.2)	80%	585	3847	20.3	108 (18.6)
T-C	541 (0.5)	84%	577	3461	19.9	108 (18.8)
TP-mf	461 (4.3)	77%	515	3621	21.0	98 (19.1)
TP-fm	471 (5.2)	81%	506	4066	23.2	97 (19.2)
TP-co	706 (1.6)	81%	768	7570	25.0	147 (19.3)

[a] Sample treatment conditions are explained in the Methods section. [b] Average number of proteins and percent coefficient of variation for two technical replicates. [c] Ratio between number of proteins identified in both technical replicates and total number of proteins identified under the corresponding treatment condition. [d] Data are reported based on proteins identified under each treatment condition with 1% FDR. [e] Percentage of integral membrane proteins as a function of total proteins for each condition.

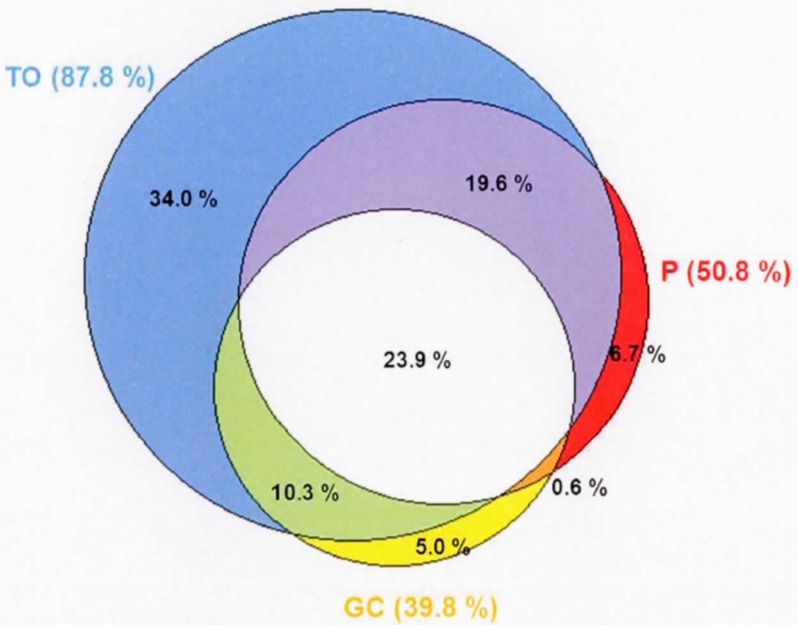


Figure 2.2 Venn diagram representing the proportion of identified proteins for single enzyme digestions as a function of the total number of proteins from trypsin (**TO**), pepsin (**P**) and Glu-C (**GC**) data sets combined

In order to examine the solubilizing power of different classes of detergents and denaturants, one reagent from each functional category was selected, while considering the conditions required for enzyme activity such as pH, temperature and critical micelle concentration (CMC). SDS represented linear-chain ionic detergents, while SDC, DDM, and FC-14 were used as steroidal ionic, non-ionic and zwitterionic detergents, respectively. Guanidinium chloride represents the chaotropic group. Methanol was used in an aqueous-organic solvent system, and trifluoroacetic acid was employed as an organic acid. Adding solubilizing agents to trypsin sometimes, but not always, improved the overall performance of proteomic analysis on rat liver microsomes. As illustrated in Table 2.1, all solubilizing agents tested enhanced the average protein sequence coverage except guanidinium chloride (**TG**) which gave a slightly higher proportion of identified IMPs, probably due to its membrane-

disruptive nature as a chaotropic agent. Guanidinium chloride is known to be a strong protein denaturant due to efficient charge-charge interactions on the protein surface (Vecchio *et al.*, 2002) and has previously been employed in the analysis of liver microsomes (Zgoda *et al.*, 2009). The best overall sequence coverage was obtained by the methanol-water solution (**TM**) and the glycosidic non-ionic detergent DDM (**TD**). However, only heated SDS (**TSH**) demonstrated a higher number of peptides and proteins comparing to trypsin alone. Methanol aqueous solution was also previously reported (Blonder *et al.*, 2004; Ye *et al.*, 2009) as an effective solubilizing agent for the analysis of microsomal membrane proteome providing a better detection of hydrophobic proteins despite decreasing trypsin activity (Mbeunkui and Goshe, 2011). The total number of identified proteins portrays the overall proteome coverage and, in this set of experiments, was best achieved with the **TSH** treatment, also yielding very good results in terms of characterization of IMPs and overall protein sequence. Therefore, heated SDS solubilization combined with trypsin digestion was found to be the best single enzyme treatment condition for proteomic analysis of rat liver microsomes. Concerning compatibility issues with MS detection, because of TCA purification and SCX fractionation steps, excess detergent is removed prior to applying samples to reversed-phase chromatography. Also, the use of an automated switch valve during the LC-MS/MS analysis ensured that residual non-volatile salts from SCX fractions did not interfere with downstream electrospray ionization.

2.4.2 Sequential Dual-Enzyme Digestion

In a second set of experiments, the idea of using two complementary digestions was tested using trypsin in combination with either pepsin or cyanogen bromide (chemical digestion). Although it was assumed that a dual digestion of rat liver microsomes with trypsin followed by a complementary enzyme (pepsin) would exhibit better overall performance, none of the combinations tested represented any

superiority over trypsin alone. It was interesting though that: (1) performing pepsin digestion before trypsin (**P-T**) digestion led to better results than the reverse order (**T-P**), especially in terms of enhanced protein sequence coverage, and (2) solubilizing the microsomes in SDS when two digestions were performed in series actually decreased the overall performance (Table 2.1). This unanticipated result could potentially be explained by interference between SDS and the acidic environment needed during pepsin digestion, or the decreased activity of pepsin in the presence of residual SDS during digestion. There is also the notion that using two enzymes in series adds complexity to the peptide mixture that is to be fractionated, and thus each fraction analyzed by LC-MS/MS is, as a result, much more complex. This can lead to increased ion suppression if too many peptides are eluting from the column at the same time, as well as a decreased ability to sequence each peptide eluting based on data-dependent triggering for MS/MS acquisition. There is most likely room for improvement in terms of results from sequential digestion samples if the resolution of the pre-fractionation or reversed-phase separation was increased to deal with the higher complexity of the sample. For this study, however, it was deemed more appropriate to keep all chromatographic steps identical between comparative treatment conditions as well as not to have extremely long LC-MS/MS run times.

2.4.3 Parallel Dual-Enzyme Digestion

Parallel dual enzyme digestions were carried out in three different configurations. In the first treatment, tryptic (**TSH**) and peptic (**P**) digestions were performed separately, and then resulting digests were mixed and fractionated together (**TP-mf**). In the second treatment, digests were fractionated separately, and then combined prior to LC-MS/MS (**TP-fm**). In the third parallel digestion strategy, trypsin and pepsin digests were fractionated and analyzed separately by LC-MS/MS,

followed by compiling mass spectrometry data from both digests and processing the data together in ProteinPilot (**TP-co**). As illustrated in Table 2.1, mixing digests after fractionation led to better results than when they were combined prior to SCX purification. This shows that the resolution of the ion-exchange separation decreases with increased complexity of the peptide digest. This step could potentially be optimized for increased resolution by either using a higher capacity chromatographic column or a longer gradient elution profile. Less proteins were identified in parallel dual enzyme digestion comparing to trypsin digestion alone, except for the condition when separate mass spectral data were co-processed from individual tryptic and peptic SCX fractions (**TP-co**), which yielded ideal results both in terms of quantity (number of proteins and peptides) and quality (sequence coverage) of data. This treatment condition is a good example of a dual enzyme digestion procedure in which two proteases act to compensate the practical limitations of each other, and thus increase the overall protein sequence coverage by generating more cleavage peptides per protein. As the results demonstrate, this treatment yielded more identified peptides by MS/MS, translating into a higher number of proteins and better overall sequence coverage. Furthermore, for the 347 proteins shared between **TSH**, **P** and **TP-co**, a significant increase in average protein sequence coverage was observed for the tryptic-peptic dual digestion (**TP-co**, 38.7%) comparing to tryptic (**TSH**, 30.7%) or peptic (**P**, 16.5%) digestions alone (see Supplemental Information S1c). In addition, 12 proteins were found to be unique to the **TP-co** condition representing the increased statistical value of combining data sets before processing with ProteinPilot versus simply combining separate **TSH** and **P** searches.

The difference in sequence coverage between trypsin and pepsin, used separately and in parallel is interesting to illustrate for protein of different hydrophobicities. Table 2.2 presents a comparison for a globular (soluble) protein, a peripheral membrane protein and an integral membrane protein in terms of % sequence coverage and number of confident peptides. Glutathione *S*-transferase Mu 2

[P08010], a globular protein, demonstrated much better results for trypsin (**TSH**) than pepsin (**P**) and since the protein is relatively hydrophilic (GRAVY < 0) and rich in tryptic cleavage sites, trypsin-pepsin parallel dual digestion (**TP-co**) did not yield better results than trypsin alone. Cytochrome P450 2D26 [P10634], which is a peripheral membrane protein (PMP) and is amphiphilic (GRAVY ≈ 0), represented a less significant increase in performance with trypsin compared to pepsin, and the parallel dual digestion procedure was more advantageous in this case. In contrast, pepsin digestion provided better results than trypsin for microsomal glutathione *S*-transferase 1 [P08011], an integral membrane protein (IMP). For this relatively hydrophobic (GRAVY > 0) membrane protein, parallel dual digestion further increased the overall digestion performance and represented a more significant complementary behavior for the two enzymes since cleaving at both tryptic and peptic sites led to higher protein sequence coverage. This comparison illustrates the intrinsic ability of the parallel dual-solubilization/digestion procedure for higher sequence coverage over the traditional single digestion, and its increased applicability for the analysis of samples enriched in membrane proteins.

Table 2.2 Comparison between sequence coverage (%) and number of confident peptides (in parentheses) for three selected proteins

Protein Name	Accession	Type	GRAVY	TSH	P	TP-co
Glutathione *S*-transferase Mu 2	P08010	Globular	-0.5	66.1% (18)	11.0% (2)	68.4% (20)
Cytochrome P450 2D26	P10634	PMP	0.0	52.6% (37)	31.8% (16)	62.6% (54)
Microsomal Glutathione *S*-transferase 1	P08011	IMP	0.2	58.1% (19)	71.6% (35)	88.4% (56)

2.4.4 Analysis of Identified Proteins

In this study, a total of 1095 proteins were identified in rat liver microsomes with a 1% global false discovery rate (FDR) out of which 213 proteins (19.5%) have been predicted to be IMPs by Phobius, a combined trans-membrane topology and signal peptide predictor. The rest are presumed to be either PMPs, other globular proteins from ER, Golgi apparatus and mitochondria matrix, or simply, vesicle-trapped cytoplasmic proteins co-purified during microsomal preparation. Phobius has been reported as the best method to distinguish membrane from non-membrane proteins in large-scale proteomics studies compared to traditional transmembrane domain (TMD) predictors (such as TMHMM, SCAMPI and SOSUI) (Krogh *et al.*, 2007; Tsirigos *et al.*, 2012). It has also been shown (Xiong, 2006) that the prediction accuracy is 94% for Phobius compared to 70% for TMHMM. Prediction results obtained from our data set were verified using corresponding UniProt annotations of identified proteins. This comparison yielded 31%, 40% and 43% false positives for TMHMM, SCAMPI and SOSUI algorithms, respectively, and only 8% for Phobius, further proving the superiority of Phobius when it comes to predicting membrane proteins.

This large data set gives a very comprehensive view of the proteome in rat liver microsomes, especially considering the restriction of 1% FDR for defining an identified protein. This analysis represents a massive increase in the overall proteome coverage compared to previous proteomic studies of rat liver microsomes (Huang *et al.*, 2011; Galeva and Altermann, 2002; Arnold *et al.*, 2004; Nisar *et al.*, 2004), where different strategies have been employed. In one study, 88 proteins were identified using a 2D-GE separation coupled to peptide mass fingerprinting (PMF) by MALDI-TOF-MS (Galeva and Altermann, 2002). Arnold *et al.* (2004) tested the use of urea (chaotrope) and ALS (RapiGest mass-compatible detergent) for identifying proteins by nanoLC-MS/MS in different samples from rat liver, including microsomal

fractions with a total of 143 microsomal proteins being identified. Also, a more targeted analysis for cytochrome P450 isoforms (CYPs) in rat liver microsomes was performed by SDS-PAGE-LC-MS/MS identifying a total of 24 CYPs (Nisar *et al.*, 2004). In a more recent publication, 391 proteins were identified (at 95% confidence) by 2D(RP-RP)-nanoLC-MS/MS (Huang *et al.*, 2011). The proteins identified in the present study were compared to a compiled list of 232 proteins from these aforementioned studies (Huang *et al.*, 2011; Galeva and Altermann, 2002; Nisar *et al.*, 2004) with a resulting overlap of 90.5%. Moreover, the 2D-LC-MS/MS method used here represented a relatively high reproducibility (71-84%) in terms of proteins identified in duplicate samples (Table 2.1). This reproducibility is mainly a consequence of the ability of the mass spectrometer to acquired MS/MS at very high speeds as well as the dynamic background subtraction IDA criteria which alleviates the problem of triggering MS/MS acquisition on background ions. Therefore, most of the proteome is reproducibly detected, even when we take into account separate digestions and SCX fractionations.

Of note, the current study utilized much higher flow rates coupled to the mass spectrometer than traditional proteomics workflows using nanoLC-MS/MS. This increased flow rate affords improved robustness and reproducibility, as well as a higher throughput. Nanoscale chromatography is considered the most sensitive approach, however, when we take into account the increased loading capacities of micro-columns as well as the improvements in ionization efficiencies at high flow rates in most modern mass spectrometers, this effect becomes less apparent. Furthermore, the ultra-high pressure system and sub-2 μm core-shell column used yielded very high chromatographic resolution, especially considering the reasonably fast gradient used for peptide elution, in comparison to several hours used in some nanoscale applications for increasing proteome coverage. This fact also contributed to increased sensitivity since peak widths (on average 8-10 seconds at half height) were decreased in comparison to longer gradients coupled to nanoscale systems.

From this compiled data set, the rat liver microsomal proteome was analyzed in terms of several physicochemical properties. First, a comparison of isoelectric point and molecular weight distribution for the 1095 identified proteins was performed (Figure 2.3). Considering the molecular mass distribution of the proteome, 63% of proteins were below 50 kDa, 29% in the range from 50 to 100 kDa, and 8% above 100 kDa. Also, a reasonably wide range of theoretical isoelectric points, from pH 4-12, was found. This offers the potential for protein-level pre-fractionation based on isoelectric focusing as a complementary fractionation step (Michel *et al.*, 2003), which could in turn further increase the resolution of the proteomic analysis. In terms of molecular weight, there is a less dramatic distribution, however, an initial protein separation could also be envisioned using size-exclusion chromatography, gel electrophoresis or, the more recently described and promising technique of Gel-Eluted Liquid Fraction Entrapment Electrophoresis (GELFrEE) (Tran and Doucette, 2008).

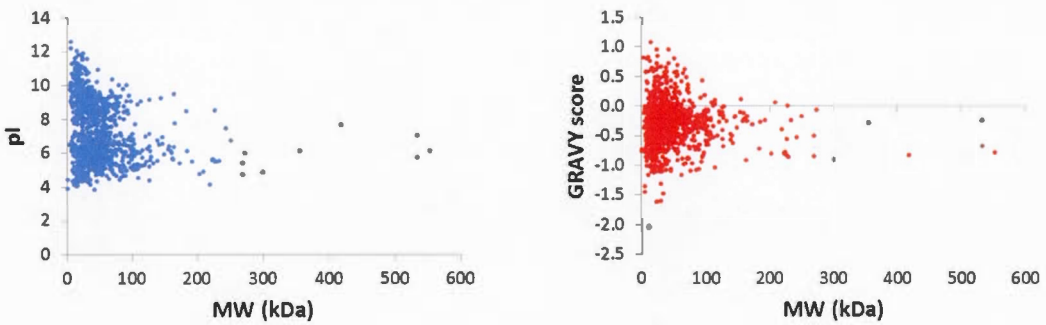


Figure 2.3 2D map of isoelectric point (left) and GRAVY scores (right) for the proteins identified in this study as a function of molecular weight

The list of identified proteins was also analyzed in terms of hydropathy (as measured by GRAVY scores). Here, the mean GRAVY score of the RLM was -0.27, with 17% of identified proteins being characterized as hydrophobic, thus supporting the predicted percentage of integral membrane proteins from Phobius (Figure 2.3).

Overall, an increase in hydrophobicity is seen for smaller proteins, a fact which has also been noted in a previous proteomic study on membrane proteins (Fischer and Poetsch, 2006). It was also observed that peptic digestion yielded IMPs with more sequence coverage of TMDs compared to trypsin, even though the latter led to a higher number of IMPs being identified. Figure 2.4 illustrates that not only were more IMPs identified by tryptic-peptic parallel dual digestion than either enzyme alone, but more TMDs were also obtained with dual digestion. It was also found that other parallel dual digestion strategies are not as efficient as the separately treated **TP-co** method regarding the identification of TMD-rich membrane proteins (Figure 2.4).

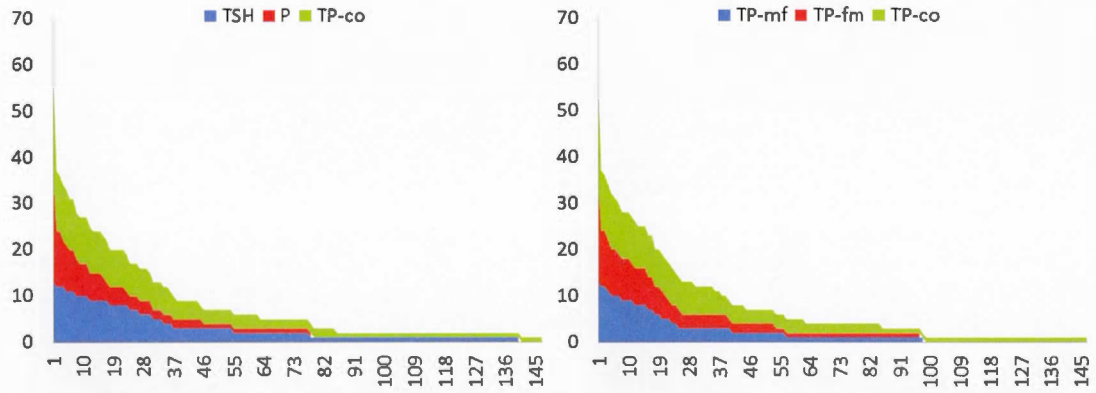


Figure 2.4 Number of transmembrane domains (TMD) of identified integral membrane proteins (IMP) under tryptic, peptic and parallel separately treated tryptic-peptic digestions (left) and under three different strategies of tryptic-peptic parallel dual digestion (right)

Other than simply stating the average sequence coverage as in Table 2.1, we have refined results of the complete data set to look at the distribution of sequence coverage throughout the proteome. This demonstrated that 36% of the proteins were identified using one peptide, 21% by 2-3 peptides, 25% by 4-10 peptides, and 18%

with over 10 peptides. It also revealed that 10% of the identified proteins had over 50% sequence coverage, 33% over 25% and 57% over 10% (Figure 2.5). Since liver microsomal fractions are often used in molecular toxicology studies, it would be ideal to have the highest possible sequence coverage if the regulation of post-translational modifications is involved in a pathway of interest.

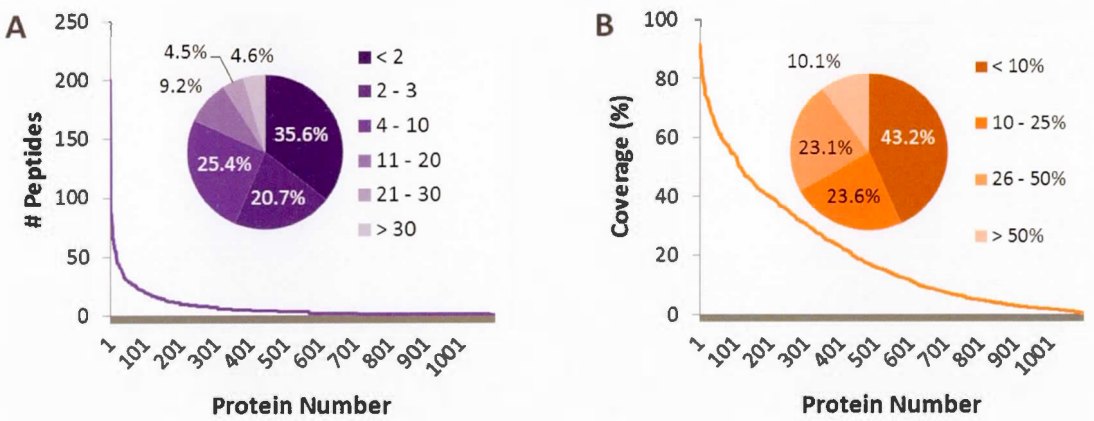


Figure 2.5 Number of confident peptides per protein (A) and overall protein sequence coverage (B) from entire data set (1095 proteins in total)

Detailed analysis was also performed using PANTHER (Protein Analysis through Evolutionary Relationships) in order to classify the identified proteins by gene ontology (GO). As shown in Figure 2.6, GO analysis classified the identified rat liver microsomal proteins into 11 groups based on their molecular functions with a major contribution representing catalytic functions (45%), followed by binding and structural functions. When classified by biological processes, 17 groups were identified with the highest proportion being involved in metabolic processes (35.2%), which makes perfect sense when we consider that liver microsomes are highly enriched in metabolic enzymes.

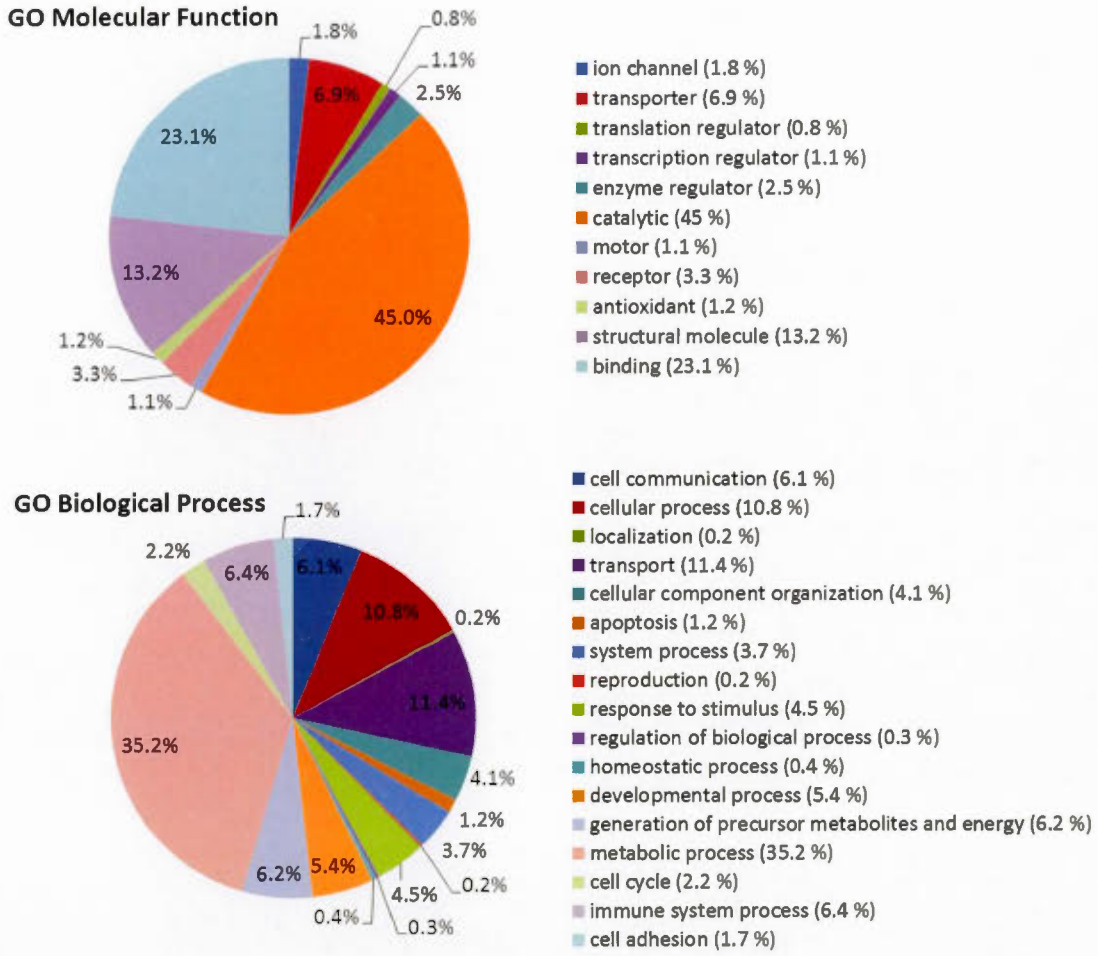


Figure 2.6 GO annotations for identified rat liver microsomal proteins based on their molecular function (above) and biological processes (below)

The subset of proteins involved in metabolic processes was analyzed for the presence of known protein-protein interactions using InnateDB. The network shown in Figure 2.7 (visualized with the Cerebral plugin of Cytoscape software) is a representation of all the known interactions that were found. These interactions are displayed within the context of subcellular localization. Since InnateDB does not incorporate information of rat interactome, human orthologs were used in this analysis. The bulk of the interactions between metabolic proteins visualized in this

figure are located within the nucleus and the cytoplasm, which includes all organellar membranes, most of which are presumed to be ER-associated. Protein interaction networks will help subsequent studies planned in our group aimed at understanding the mechanism of xenobiotic-driven toxicity due to irreversible modification of proteins. Modification of a key protein may affect many other proteins via interaction networks and in turn affect the overall response to covalent binding. A similar analysis as was performed in Figure 2.7 could be carried out with other protein subsets or across the whole proteome by submitting accession numbers of interest to InnateDB in order to have a closer look on liver microsomes from other perspectives as well.

In fulfilling such a comprehensive proteome analysis of rat liver microsomes, the information contained in this data set could support further studies of biological processes applied to drug discovery and toxicological research. A list of identified proteins found in each of the tested treatment conditions, as well as a compiled comprehensive list of all identified proteins, is provided as supporting information S1. Supplemental information is also provided for individual peptides which contributed to the identification of each protein (S2). This list of peptides could be especially useful for developing targeted quantitative assays, based on multiple-reaction monitoring (MRM) detection.

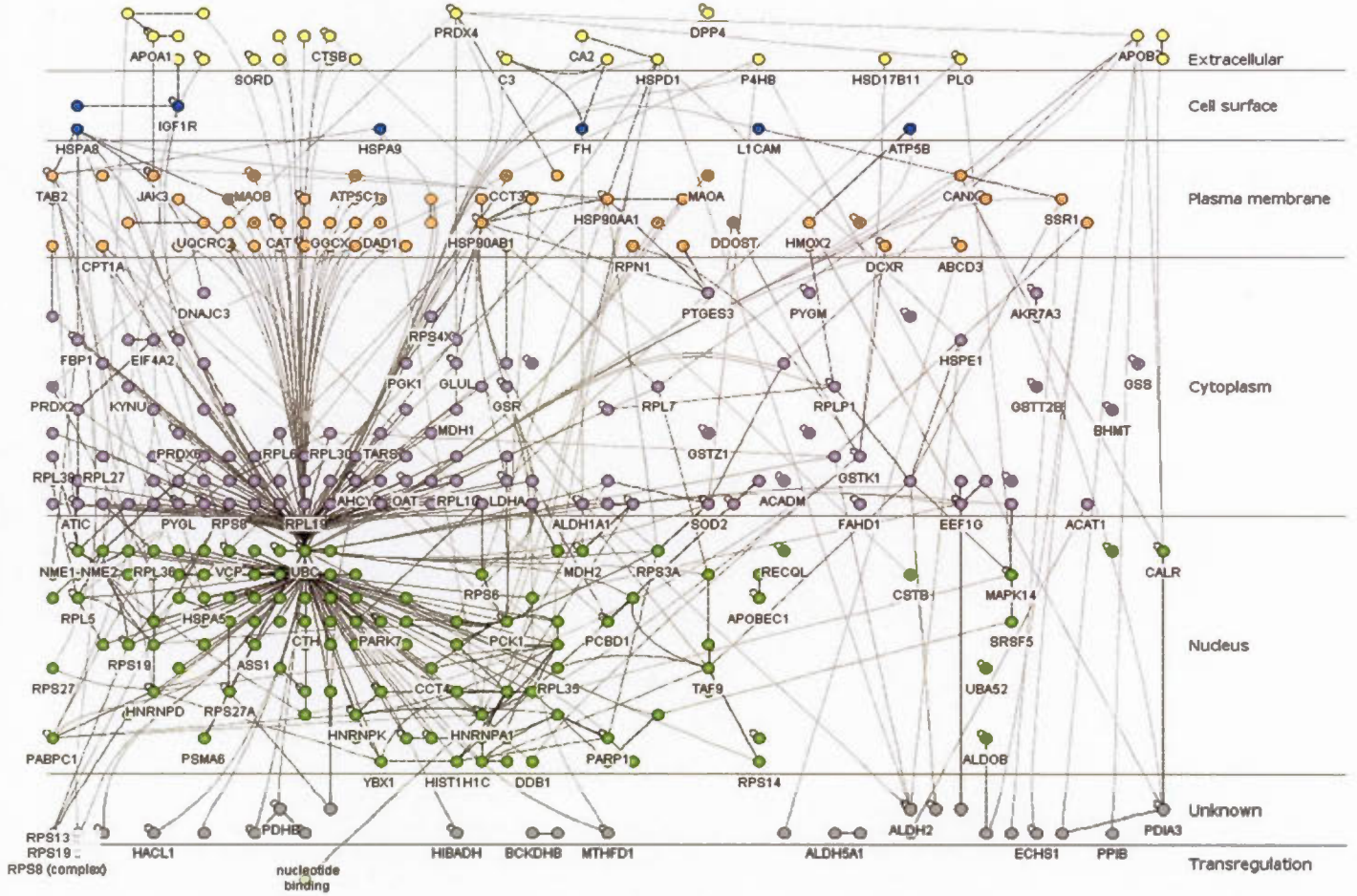


Figure 2.7 A subset human analogs of identified proteins reported to be involved in metabolic processes analyzed for the presence of known human protein-protein interactions using InnateDB

2.5 Conclusions

The comparative analysis carried out in this study characterized an efficient strategy for bottom-up proteomic analysis of rat liver microsomes. Examining eighteen different digestion conditions allowed several conclusions to be made. For single enzyme digestions, trypsin gave better results than pepsin or Glu-C. Among the seven tested solubilizing agents, SDS at elevated temperature represented the best overall performance and was easily removed by SCX fractionation prior to LC-MS/MS analysis. Tryptic-peptic sequential dual enzyme digestions did not increase the quality of results compared to using trypsin alone, however in terms of digestion sequence, using pepsin prior to trypsin was the better choice. Co-processing LC-MS/MS data of separately analysed tryptic and peptic digests yielded the best results with the highest number of peptides and identified proteins. Therefore, co-processed parallel enzyme digestions using trypsin, using SDS and heat to aid solubilization, in combination with pepsin led to an optimized strategy for comprehensive proteomic analysis of rat liver microsomes. This method provided the highest number of proteins with the best protein sequence coverage compared to all other treatments assessed in this work. Higher sequence coverage allows more in-depth analysis of the proteome, especially for characterizing protein modifications which are of great interest in drug discovery and toxicology applications. Also, a good characterization of integral membrane proteins was observed using this optimized procedure. This comprehensive study of the RLM proteome will be useful in various future applications, including drug discovery and toxicological studies.

2.6 Supporting Information

(S1) Protein database: (a) list of proteins identified for each treatment condition (b) false discovery rate reports for each condition (c) detailed analysis of overlap in proteins identified in **TSH**, **P** and **TP-co** conditions (d) Integrated list for all 1095 identified proteins. (S2) Peptide database: compiled list of peptides identified for each protein under each treatment condition.

CHAPTER THREE

MULTIDIMENSIONAL LC-MS/MS ANALYSIS OF LIVER PROTEINS IN RAT, MOUSE AND HUMAN MICROSOMAL AND S9 FRACTIONS

Makan Golizeh, Christina Schneider, Leanne B. Ohlund and Lekha Sleno

Published in **EuPA Open Proteomics 2015, 6, 16-27**. Supplementary data available online
doi:10.1016/j.euprot.2015.01.003.

In this chapter, different combinations of orthogonal LC techniques were tested to enhance the number and sequence coverage of identified proteins in rat liver microsomal samples. To this end, protein- and peptide-level ion exchange fractionation was coupled to RP-LC-MS/MS and compared. A peptide-level fractionation was then employed for the analysis of rat, mouse and human liver S9 and microsomal fractions. The results of this analysis provided information for cross-species studies. The improved workflow developed here was also employed in Chapter 5 for the identification of adducted rat liver microsomal proteins.

Makan Golizeh, Christina Schneider, Leanne Ohlund and Lekha Sleno are co-authors of this article. Makan Golizeh conducted the literature survey, prepared experimental protocols, conducted the experiments, processed the data, and prepared the original manuscript. Christina Schneider, a summer intern, actively participated in the experimental part, data processing, and preparation of the results. Leanne Ohlund provided assistance with the literature survey and development of the CATWAX-LC method. Professor Lekha Sleno supported and supervised the project, verified data analysis and interpretation of the results, revised and finalized the manuscript.

3.1 Abstract

Liver plays a key role in metabolism and detoxification, therefore analysis of its proteome is relevant for toxicology and drug discovery studies. To optimize for high proteome coverage, protein and peptide-level ion exchange fractionation were assessed using rat liver microsomes and S9 fractions. 2D-(SCX-RP)LC-MS/MS analysis with peptide fractionation was subsequently employed for rat, mouse and human samples, yielding between 1400 and 1939 identified proteins, 58% of which were shared between species, and with relatively high sequence coverage. This rich dataset is specifically interesting for the toxicology community, and could serve as an excellent source for targeted assay development.

3.2 Introduction

Multidimensional liquid chromatography (MDLC) improves the separation of highly complex mixtures, and has been applied to proteomics, lipidomics and the analysis of natural compounds (Holčapek *et al.*, 2012). In both qualitative and quantitative proteomics, MDLC has proven to be a powerful tool to increase the coverage and sensitivity of protein profiling, as well as improving the accuracy and reproducibility of quantitative analysis (Wu *et al.*, 2012; Xie *et al.*, 2012). Generally, compared to a single LC, combining multiple chromatographic separations will enhance resolving power, limit of detection and dynamic range by significantly increasing peak capacity (Di Palma *et al.*, 2012). Common LC techniques used in MDLC often include high-pH and low-pH RP, anion or cation exchange, size-exclusion chromatography as well as hydrophilic interaction chromatography (HILIC) and affinity chromatography (Nice *et al.*, 2007). Orthogonality of separation is an important parameter for achieving a high peak capacity, thus most common

examples include high-low pH RP-RP, strong cation exchange (SCX)-RP and HILIC-RP for expanding proteome coverage (Gilar *et al.*, 2005).

Ion-exchange chromatography is often used as a first dimension of separation. In principle, there are two modes of ion exchange separation: cation and anion exchange. Combining SCX with RP chromatography offers high orthogonality, as SCX has a high loading capacity while RP can achieve high resolution separations amenable to LC-MS analysis and both involve different mechanisms (Peng *et al.*, 2003). It is also possible to combine two complementary ion (cation and anion) exchange columns either in tandem or as part of a mixed-bed column for increased separation efficiency. Havugimana *et al.* highlighted (2007) an improved number of identified proteins and higher quality of proteomics data using a "dual-column" approach with weak anion exchange (WAX) column coupled to a moderate cation exchange column for the analysis of mouse heart cytosol. Motoyama *et al.* published, in the same year, a study with increased number of identified proteins in yeast using a mixed-bed weak cation/weak anion exchange (CATWAX) column. The use of a 3D-LC-MS approach incorporating CATWAX and SCX for protein and peptide fractionation was also reported by Zhang *et al.* (2012) to facilitate in-depth protein identification in mouse mammary tumor 4T1 cell lysate.

Liver plays a crucial role in the metabolism of xenobiotics through numerous enzymes involved in detoxification. A significant portion of these enzymes can be enriched in the microsomal fraction during subcellular fractionation (Van Summeren *et al.*, 2012) and therefore liver microsomes are often used in drug metabolism studies. Microsomes, in general, are rich sources of membrane proteins from endoplasmic reticulum, plasma membrane, mitochondria, and Golgi apparatus (Mathias *et al.*, 2011). Integral membrane proteins (IMPs) represent the most pharmaceutically useful class of receptors (Imming *et al.*, 2006), the targets of 70–80% of all known drugs (Mathias *et al.*, 2011), having strong implications in cell

survival (Helbig *et al.*, 2010). However, they are often tightly bound or physically integrated into the membrane and thus pose technical difficulties for proteomic analysis (Gilmore and Washburn, 2010). Liver S9 fraction, a mixture of microsomal and cytosolic fractions, is another important sample to study as it and is also often employed in toxicological and drug metabolism studies (Plant, 2004). Since an important goal in proteomics research aims at determining post-translational modifications (including covalent binding from reactive endogenous/exogenous species) (Zhou *et al.*, 2005), an ideal method would achieve very high sequence coverage of all potentially-targeted proteins.

We previously reported (Golizeh and Sleno, 2013) an approach for the proteomic analysis of rat liver microsomes examining different combinations of proteases and solubilizing agents using single digestion, serial dual digestion and parallel dual digestion workflows. An SDS-assisted parallel tryptic-peptic dual digestion method exhibited the highest proteome coverage with 768 proteins identified at 1% global false discovery rate (25% average protein sequence coverage) with a high proportion (19.3%) of integral membrane proteins. To further enhance proteome coverage, the present study compares several MDLC-based separations of rat liver microsomal and S9 fractions incorporating both protein-level and peptide-level fractionation. Four strategies were devised to optimize the proteomic analysis of rat liver samples combining ion exchange with reversed-phase chromatography coupled to high resolution quadrupole-time-of-flight (QqTOF) mass spectrometry. The approach with the best overall performance was then selected to carry out a cross-species proteomics comparison of rat, mouse and human liver fractions.

The main goal of this work was to improve the proteome coverage for samples of specific interest to drug metabolism and toxicology research, relatively rich in membrane proteins, where achieving high sequence coverage is particularly challenging. The increased coverage of individual proteins enhances the confidence

of their detection with a higher number of peptides per protein, therefore yielding a better potential for accurate protein quantitation. For instance, this study is particularly useful for applications involving the absolute quantitation and expression profiling of cytochrome P450 and UDP-glucuronosyltransferase enzymes (Langenfeld *et al.*, 2009; Bucher *et al.*, 2011; Schaefer *et al.*, 2012). Moreover, when specific protein modifications are studied, higher protein coverage would increase the chances of seeing such modifications in complex samples.

3.3 Experimental

3.3.1 Materials

Male Sprague–Dawley rat, male ICR/CD-1 mouse, and human (M-50 donor) pooled liver microsomes (20 mg/ml protein) were purchased from Celsis In Vitro Technologies (Baltimore, MD). Rat liver S9 fractions (37.5 mg/ml protein) were from Moltox (Boone, NC), while male CD-1 mouse and human liver S9 fractions (20 mg/ml protein) were purchased from BD Biosciences (Franklin Lakes, NJ). Sequencing-grade modified trypsin was obtained from Promega (Madison, WI). Porcine gastric mucosa pepsin and all other chemical reagents were purchased from Sigma-Aldrich (St. Louis, MO). HPLC-grade ACN, methanol and isopropanol were from Caledon (Georgetown, ON), and ultra-pure water was from a Millipore Synergy UV system (Billerica, MA).

3.3.2 Sample Digestion

Liver microsomes or S9 fractions (0.5-0.6 mg protein) were solubilized with a 2% SDS solution (1:1, v/v), heated at 95°C for 3 min, and diluted with 0.1 M

ammonium bicarbonate (pH 8.5) to 200 µl. Reductive alkylation was performed using DTT (2.5 mM, 10 min, 25°C) and iodoacetamide (5 mM, 30 min, 37°C, dark). Samples were diluted with 150 µl of 0.1 M ammonium bicarbonate (pH 8.5) for trypsin digestions or 0.2% TFA in 20% methanol for pepsin digestions, and incubated at a 1:50 (w/w) enzyme : protein ratio for 18 h, at 37°C for trypsin and 25°C for pepsin with an additional 15 µl of 10% TFA prior to adding pepsin samples to maintain the required acidic digestion conditions (pH 1.5-2). CATWAX protein fractions were reconstituted in 150 µl of 0.1 M ammonium bicarbonate (pH 8.5), denatured and digested using 5 µg of the enzyme. Digestion was quenched by adding 50 µl of 1% formic acid for trypsin or 1% ammonium hydroxide for pepsin. Samples were then diluted with 500 µl H_2O prior to solid-phase extraction (SPE) on a 1 cc (30 mg) OASIS HLB cartridge (Waters, Milford, MA), eluting with 100% methanol (1 ml). Resulting samples were evaporated to dryness under vacuum (Thermo Fisher Scientific Universal Vacuum System, Asheville, NC) and stored at −30°C.

3.3.3 Protein Fractionation

Solubilized protein samples (1.4 mg total) were diluted with buffer A (see below) to have a total volume of 280 µl (5 mg/ml), sonicated (1 min) and then filtered using a Costar Spin-X 0.45 µm cellulose acetate centrifuge tube filter (Corning, Corning, NY) prior to injection (100 µl, 0.5 mg protein) onto a PolyCATWAX 200 × 2.1 mm column with 5 µm (1000 Å) particles (PolyLC, Columbia, MD) using an Agilent 1200 series HPLC (Agilent Technologies, Palo Alto, CA) equipped with a binary pump, degasser, diode array detector and fraction collector. Mixed-bed weak cation/weak anion exchange (WCX/WAX or CATWAX) fractionation was performed at a flow rate of 200 µl/min using a gradient of 100% A held for 1 min, up to 8% B at 8 min, 85% B at 9.5 min, then 100% B at 12.5 min, and held at 100% B for an additional 17.5 min, where buffers A and B were 10 mM and 800 mM

ammonium acetate in 20% acetonitrile (pH 7.2), respectively. The stand-alone weak cation exchange (WCX) and weak anion exchange (WAX) separations were performed on PolyCAT and PolyWAX 100 × 2.1 mm columns with 5 μm (1000 Å) particles (PolyLC, Columbia, MD), respectively. The WCX and WAX columns were also used in serial WCX-WAX and WAX-WCX configurations for dual-column weak cation/weak anion exchange experiments. In all cases, the UV absorbance was monitored at 220 and 280 nm. From the CATWAX separation, six fractions were collected over 22 min. The fractions were collected at the following intervals: 0-2, 2-4, 4-6, 6-14, 14-18 and 18-22 min, which were then evaporated to dryness under vacuum and kept at −30°C prior to digestion.

3.3.4 Peptide Fractionation

Protein digests were reconstituted in buffer A (120 μl, see below), sonicated (10 min), and centrifuged (5 min, 14,000 g) prior to injection (100 μl, 0.5 mg protein) onto a Zorbax 300-SCX 150 × 2.1 mm column with 5 μm (300 Å) particles (Agilent Technologies, Palo Alto, CA) using the same HPLC system as for protein fractionation. SCX fractionation was performed at a flow rate of 250 μl/min with a gradient of 0–50% B in 15 min, up to 100% B at 25 min, then held for an additional 5 min at 100% B, where buffers A and B were 10 mM potassium dihydrogen phosphate in 25% acetonitrile (pH 2.75), and 1 M potassium chloride in buffer A (pH 2.75), respectively. UV absorbance was monitored at 220 and 280 nm. For trypsin samples, 3 min (0.75 ml) fractions were aliquoted into 1.5 ml tubes between 1.5 and 19.5 min, while for pepsin, 4 min (1.0 ml) fractions were collected between 1.5 and 25.5 min. Fractions were evaporated to dryness under vacuum and kept at −30°C.

3.3.5 RP-UHPLC-MS/MS Analysis of Peptide Fractions

Dried samples were reconstituted in 10% acetonitrile (100 µl) and injected (20 µl) onto an Aeris PEPTIDE XB-C_{18} 100 × 2.1 mm column, with solid core 1.7 µm particles (100 Å) (Phenomenex, Torrance, CA) on a Nexera UHPLC system (Shimadzu, Columbia, MD) with water (A) and acetonitrile (B), both containing 0.1% formic acid, at a flow rate of 300 µl/min (40 °C). The gradient started at 5% B, held for 2 min, and was increased linearly to 30% B at 24 min, to 50% B at 26 min, then to 85% B at 26.5 min. MS and MS/MS spectra were collected on a high-resolution TripleTOF 5600 mass spectrometer (AB Sciex, Concord, ON) equipped with a DuoSpray ion source in positive ion mode. The instrument performed a survey TOF-MS acquisition from *m/z* 140–1250 (250 ms accumulation time), followed by MS/MS on the 15 most intense precursor ions from *m/z* 250–1250 (excluded for 20 seconds after two occurrences) using information-dependent acquisition (IDA) with dynamic background subtraction. Each MS/MS acquisition had an accumulation time of 50 ms and collision energy of 30 ± 10 V. The total cycle time was 1.05 s.

3.3.6 Data Treatment

MS/MS files from each workflow were combined and searched against the UniProt protein database (release date 26/06/2013) by ProteinPilot software (version 4.1) for the specified species (rat, mouse, or human) using Paragon algorithm (Shilov *et al.*, 2007), including false discovery rate (FDR) analysis and detection protein threshold of unused ProtScore > 0.05 (confidence >10%). The search was performed for +2 to +4 charge states and MS tolerance was 0.05 Da on precursor ions and 0.1 Da on fragments. All duplicates were first processed alone, then together and finally tryptic and peptic digest for each sample were co-processed together. All strategies were then combined into a "master" file to represent the total number of

proteins and peptides identified from all methods. Proteins were identified with a 1% global false discovery rate (FDR) using a target-decoy database search algorithm (Tang *et al.*, 2008).

The list of UniProt accession numbers from identified proteins was uploaded to NCBI Batch Entrez to obtain the batch FASTA file, which was subsequently submitted to ExPASy for determination of isoelectric point and monoisotopic molecular weight and to Phobius (Krogh *et al.*, 2007) for prediction of IMPs, based on having at least one transmembrane domain (TMD $\geq$ 1). GRAVY Calculator was used to calculate grand average of hydropathy (GRAVY) scores. The list of accession numbers was also uploaded to PANTHER (Thomas *et al.*, 2003) for gene ontology (GO) classification. Venn diagrams were created by Venny interactive Venn diagram plotter (BioinfoGP) (Oliveros, 2007) while Clustal Omega (Sievers *et al.*, 2014) was used for multiple sequence alignment (MSA) analysis of the proteins unique to each dataset.

3.4 Results and Discussion

Different sample fractionation methods were compared to achieve high proteome coverage for the analysis of rat liver microsomal (RLM) and S9 fractions (RLS). Several combinations of separation techniques including solid-phase extraction (SPE), peptide-level ion exchange fractionation, and protein-level ion exchange fractionation were evaluated (Figure 3.1). Each workflow was performed on pepsin and trypsin digested samples. The method with the best performance was then applied to the analysis of rat, mouse and human liver microsomal and S9 fractions. This cross-species comparison unveiled large overlaps between the datasets obtained and further analysis of the results shed light on possible orthologs between the species.

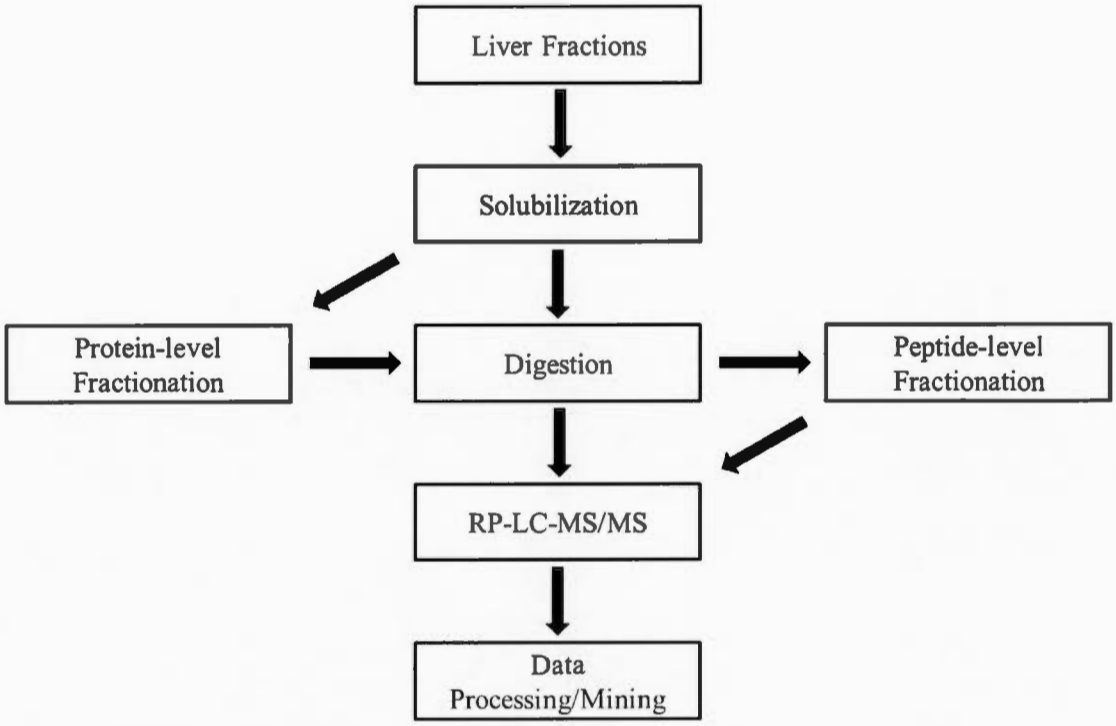


Figure 3.1 MDLC-MS/MS proteomic analysis workflow

3.4.1 Method Optimization

For protein solubilization, SDS is often used, however, it is known to cause technical difficulties, interfering with downstream MS analysis, thus needs to be effectively removed prior to LC-MS (Griffin and Schnitzer, 2011). In previous work, SDS-assisted tryptic digestion yielded better results than several other solubilizing agents when peptide-level ion exchange fractionation was performed prior to LC-MS analysis (Golizeh and Sleno, 2013). The current study also tested the effect of SDS on pepsin digestion efficiency with and without detergent removal by TCA precipitation. The SDS-solubilized peptic digestion without TCA precipitation yielded the highest number of identified proteins at a 1% final SDS concentration (in 0.1% TFA, 10%

methanol) (data not shown). During the SPE clean-up step, different elution conditions were tested on peptic digests (from RLM and RLS) using OASIS HLB cartridges (Waters, Milford, MA). Elution with 100% methanol yielded the highest number of identified peptides and proteins at 1% FDR, followed by a 60:40 mixture of ACN/isopropanol, and 100% ACN exhibited the least favorable results (unpublished data).

For LC-MS/MS acquisition, elution gradients and MS/MS parameters were assessed for the analysis of RLM digests (after SPE clean-up). The LC method employed in previous work (Golizeh and Sleno, 2013) was further optimized by varying flow rates and gradient conditions slightly, with the goal of keeping a reasonable throughput (30-40 min per run). A flow rate of 300 µl/min offered better separation efficiency compared to 250 and 350 µl/min using the solid-core 1.7µm C_{18} column. Based on the total ion chromatogram (TIC) of the TOF-MS survey scan; the gradient was modified to 5% B held for 2 min, gradually increased to 30% B at 24 min, to 50% B at 26 min, then to 85% B at 26.5 min and held for 2 min. These LC conditions yielded extremely reproducible retention times between samples over the course of this study. In terms of IDA criteria, the maximum number of candidate ions for the MS/MS dependent scans per cycle was varied between 5 and 20 ions with an accumulation time ranging from 50 to 200 ms for each MS/MS. The intensity threshold of the IDA candidate ions was also varied from 200 to 500 cps and isotope exclusion was tested at 2 or 3 Da. Moreover, MS/MS acquisition of the precursor ion was excluded after two MS/MS scans for either 15, 20 or 30 seconds. Optimization of the IDA-MS/MS parameters resulted in 15 MS/MS with 50 ms accumulation time each, a 3 Da isotope exclusion and 20 s dynamic exclusion for selected precursor ions. This led to a significant improvement of the number of identified peptides, proteins and IMPs as well as higher average protein sequence % coverage and shorter total cycle time (summarized in Table 3.1).

Table 3.1 Improvements achieved by optimization of the IDA-MS/MS parameters

	Initial conditions [a]	Optimized conditions [b]	Δ
# Peptides (1% FDR)	**7570**	**17566**	**132 %**
# Proteins (1% FDR)	**768**	**1120**	**46 %**
Avg. Seq. % Coverage	**25.0**	**29.9**	**20 %**
# IMP	**147**	**217**	**48 %**
Acquisition speed	**1.30 s/cycle**	**1.05 s/cycle**	**-19 %**

[a] 10 MS/MS scans (100 ms accumulation time) without dynamic exclusion.

[b] 15 MS/MS scans (50 ms accumulation time) with 20 s dynamic exclusion.

3.4.2 2D-LC-MS/MS Using Protein-Level Fractionation

For protein pre-fractionation of microsomal and S9 fractions, five workflows were examined in this study, incorporating WCX and WAX in different combinations. Single WCX and WAX runs were compared to tandem WCX-WAX and WAX-WCX using two columns in series, as well as mixed-bed WCX/WAX (CATWAX) using identical gradient conditions. Judging by the LC-UV traces, single ion-exchange did not yield satisfactory separations, though WAX did perform better than WCX. Tandem combinations of WCX-WAX and WAX-WCX improved the separation, although CATWAX provided more resolution due to the potential of retaining both negatively and positively charged proteins over the single column in a homogeneous manner (Figure 3.2).

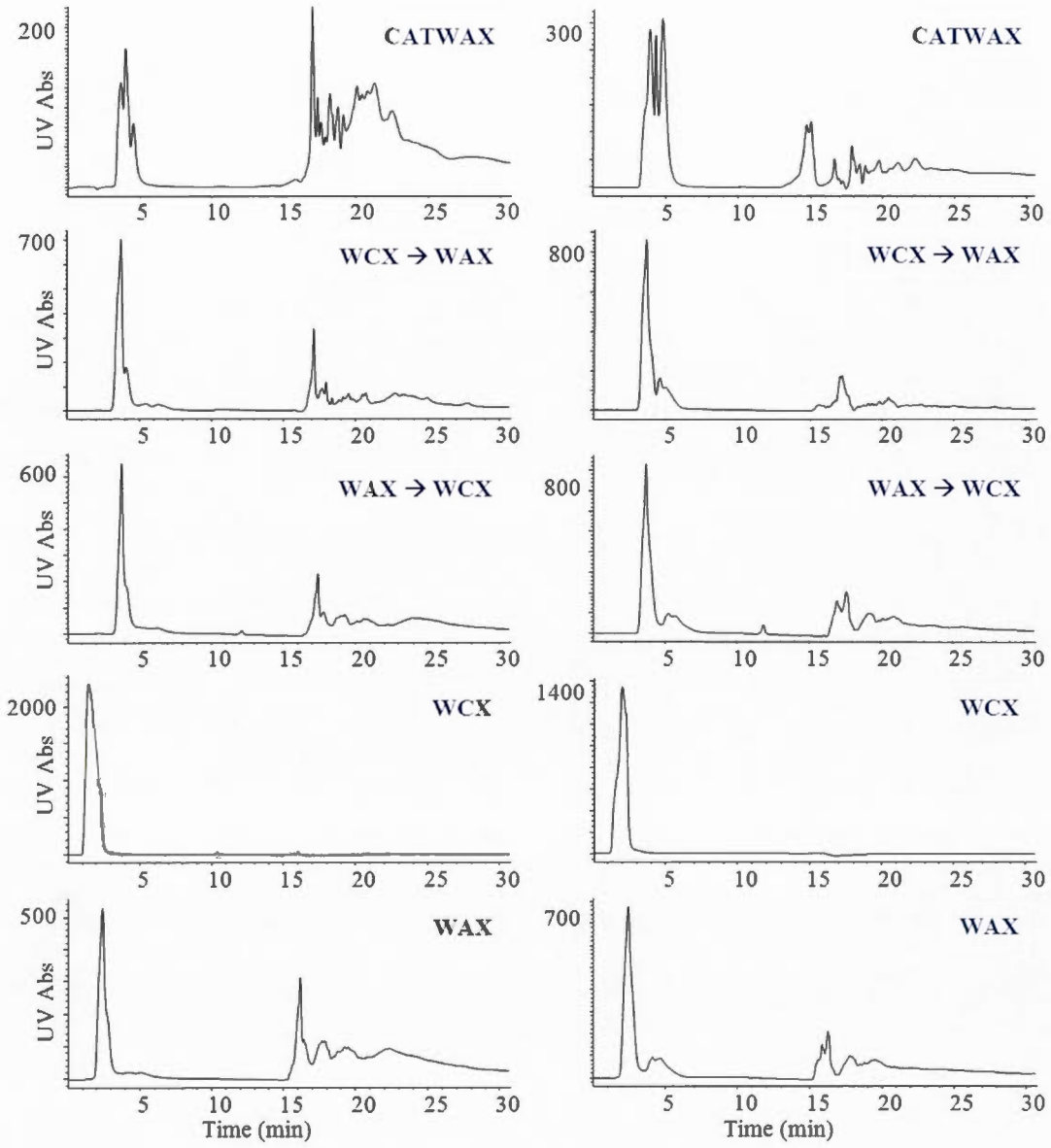


Figure 3.2 LC-UV traces at 220 nm for protein-level fractionation from mixed-bed WCX/WAX (CATWAX), tandem dual-column WCX-WAX and WAX-WCX, stand-alone WCX and WAX separations for RLM (left) and RLS (right). The y-axis (signal intensity) has been magnified 5x for the region eluting from 10-30 min for clarity.

3.4.3 2D-LC-MS/MS Using Peptide-Level Fractionation

LC-UV traces from SCX fractionation of the microsomal and S9 digested samples demonstrated a separation pattern similar to that of the protein-level fractionation (two elution zones). Tryptic digests gave a more significant UV absorption in the second portion than the earlier elution zone at 2–5 min, compared to pepsin, which seemed to have a wider elution profile overall. This is attributed to the tendency of trypsin to cleave after lysine and arginine residues resulting in more basic peptides strongly binding to the aromatic sulfonic acid groups of the SCX column. The wider spread of elution for peptic digests was attributed to more variety possible in terms of basicity as pepsin cuts at non-charged (hydrophobic) amino acids (Figure 3.3).

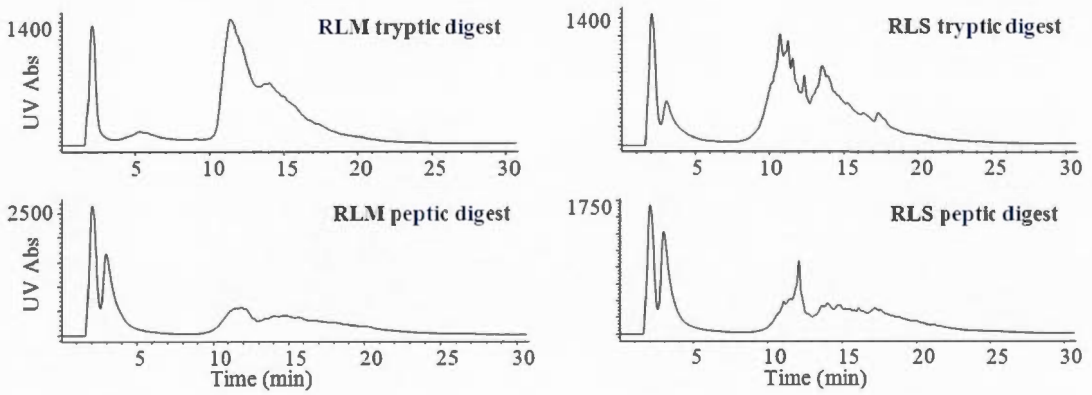


Figure 3.3 LC-UV traces at 220 nm from the SCX peptide-level fractionation of RLM (left) and RLS (right) samples digested by trypsin (top) or pepsin (bottom)

Previous work using 2D(SCX-RP)-LC-MS/MS for rat liver microsomes (Golizeh and Sleno, 2013) incorporated a protein precipitation step prior to digestion to remove membrane-associated impurities as well as excess reagents which may

hinder digestion or affect SCX separation efficiency (Westermeier and Naven, 2002). This precipitation step was replaced by a SPE clean-up step prior to SCX fractionation, to achieve good separation efficiency.

3.4.4 3D-LC-MS/MS (Combined Protein- and Peptide-Level Fractionation)

LC-MS/MS analysis with multiple dimensions of pre-fractionation has proven useful in proteomics research (Wu *et al.*, 2012), with reports of improved protein/peptide identification (Boichenko *et al.*, 2013; Li *et al.*, 2013; Rodríguez *et al.*, 2012), higher peptide selectivity (Betancourt *et al.*, 2013) and more efficient detection of modifications (McClintock *et al.*, 2013; Tousi *et al.*, 2012). However, most of these have combined successive LC steps at the peptide level. In this study, the 3D-LC-MS/MS approach tested combined protein-level (CATWAX) with dual peptide-level (SCX-RP) separations. Unfortunately, the results showed poor performance compared to both regular 1D (RP) and the 2D (SCX-RP) methods. All protein-level fractions, when subjected to peptide fractionation, yielded identical chromatographic traces. Several sample treatment workflows were tested incorporating SDS-assisted solubilization or acetone/TCA protein precipitation, however, no significant improvement was observed.

3.4.5 Tryptic versus Peptic Digestion

A parallel tryptic-peptic dual digestion was performed for the analysis of rat liver microsomes for increased proteome coverage due to the complementary cleavage sites between the two enzymes (Golizeh and Sleno, 2013). Trypsin normally cleaves at basic amino acid residues such as arginine and lysine which are more abundant in water-soluble proteins. Pepsin, on the other hand, targets residues with

hydrophobic and preferably aromatic side chains such as leucine, phenylalanine, tryptophan, and tyrosine (Dunn, 2002), likely found in less water-soluble regions, such as membrane proteins. Microsomes (and S9 fractions) are rich sources of membrane proteins and therefore a combined tryptic-peptic digestion enhances the proteome coverage as pepsin cleaves regions embedded inside (or associated with) the membrane while trypsin cuts more solvent-exposed regions. Table 3.2 summarizes the results obtained from trypsin and pepsin digestions using four fractionation strategies. As expected, trypsin led to a higher number of identified proteins and peptides, while pepsin often yielded an increased %IMP. Moreover, replicate analysis showed that, overall, trypsin samples had better reproducibility compared to pepsin samples (see Table 3.3). It was also observed that 23–51% of the identified proteins were shared between trypsin and pepsin, 37–72% were unique to trypsin, and 4–12% were unique to peptic digestion. The pepsin-unique proteins included more IMPs (up to 40%) compared to trypsin-specific proteins (up to 20%). More details on the overlap analysis can be found in Table 3.4.

Table 3.2 Comparative results from the analysis of rat liver microsomes (RLM) and S9 fractions (RLS) using the four fractionation approaches (trypsin, pepsin, and the two enzymes combined)

Sample [a]	Approach	Digestion	Proteins [b]	Peptides [b]	Avg. sequence coverage (%)	IMP (%IMP) [c]
RLM	1D	trypsin	543	4643	22.2	96 (17.7)
		pepsin	390	2453	16.0	90 (23.1)
		combined	586	7049	27.1	107 (18.3)
	2D protein (CATWAX)	trypsin	286	1384	16.1	23 (8.0)
		pepsin	138	848	20.6	7 (5.1)
		combined	297	2188	19.8	20 (6.7)
	2D peptide (SCX)	trypsin	1089	12111	24.8	217 (19.9)
		pepsin	610	5251	20.9	134 (22.0)
		combined	1120	17566	29.9	212 (18.9)
	3D	trypsin	365	1493	15.0	39 (10.7)
		pepsin	177	737	13.5	20 (11.3)
		combined	370	2109	16.8	45 (12.2)
RLS	1D	trypsin	521	4721	22.1	51 (9.8)
		pepsin	320	1817	13.4	33 (10.3)
		combined	569	6497	24.0	52 (9.1)
	2D protein (CATWAX)	trypsin	253	1190	17.8	13 (5.1)
		pepsin	74	541	24.1	2 (2.7)
		combined	260	1707	20.1	10 (3.8)
	2D peptide (SCX)	trypsin	1015	11453	25.0	101 (10.0)
		pepsin	502	4291	18.4	51 (10.2)
		combined	1037	15905	28.4	102 (9.8)
	3D	trypsin	175	676	14.4	15 (8.0)
		pepsin	67	325	17.5	4 (6.0)
		combined	174	975	15.8	14 (8.0)

[a] Data combined from duplicate samples

[b] Total number of proteins/peptides identified (from duplicate samples) with 1% FDR

[c] Percentage of integral membrane proteins identified for each condition

Table 3.3 Replicate analysis of tryptic and peptic digests for each LC-MS based approach

Sample	Approach	Digestion	Avg. # proteins	Reproducibility*
RLM	1D	trypsin	492	77%
		pepsin	343	70%
	2D protein (CATWAX)	trypsin	223	52%
		pepsin	108	50%
	2D peptide (SCX)	trypsin	990	78%
		pepsin	567	76%
	3D	trypsin	316	70%
		pepsin	138	55%
RLS	1D	trypsin	463	74%
		pepsin	255	58%
	2D protein (CATWAX)	trypsin	192	50%
		pepsin	59	54%
	2D peptide (SCX)	trypsin	934	80%
		pepsin	415	62%
	3D	trypsin	143	57%
		pepsin	57	51%

* Ratio between number of proteins identified in technical replicates and total number of proteins identified in corresponding treatment condition

Table 3.4 Number and % identified proteins and IMPs unique to and shared between tryptic and peptic digestions

Sample	Approach	unique to trypsin		unique to pepsin		trypsin ∩ pepsin	
		Proteins (% total)	IMP (% IMP)	Proteins (% total)	IMP (% IMP)	Proteins (% total)	IMP (% IMP)
RLM	1D	227 (36.8)	35 (15.4)	74 (12.0)	29 (38.7)	316 (51.2)	61 (19.3)
	2D protein (CATWAX)	179 (56.5)	19 (10.6)	31 (9.8)	3 (9.7)	107 (33.8)	4 (3.7)
	2D peptide (SCX)	529 (46.4)	103 (19.5)	50 (4.4)	20 (40.0)	560 (49.2)	113 (20.2)
	3D	234 (56.9)	29 (12.4)	46 (11.2)	10 (21.7)	131 (31.9)	10 (7.6)
RLS	1D	268 (45.6)	28 (10.4)	67 (11.4)	10 (14.9)	253 (43.0)	23 (9.1)
	2D protein (CATWAX)	193 (72.3)	12 (6.2)	14 (5.2)	1 (7.1)	60 (22.5)	1 (1.7)
	2D peptide (SCX)	553 (52.4)	63 (11.4)	40 (3.8)	13 (32.5)	462 (43.8)	38 (8.2)
	3D	129 (65.8)	14 (10.9)	21 (10.7)	4 (19.0)	46 (23.5)	0 (0.0)

3.4.6 Comparison of the Four Workflows

The results from 1D-, 2D- and 3D-LC-MS/MS approaches showed significantly superior performance of the 2D-(SCX-RP) workflow. Based on previous research (Motoyama *et al.*, 2007; Zhang *et al.*, 2012), it was expected that the best performance would be seen with the 3D strategy, combining fractionation at both protein and peptide levels. Nevertheless, the 2D-(CATWAX-RP) and 3D-(CATWAX-SCX-RP) methods led to less satisfactory results even compared to the

regular 1D-RP (Table 3.2). This is most likely related to sample loss related to protein fractionation. Proteins identified exclusively with sample preparation workflows involving protein fractionation were compared between RLM (94 proteins) and RLS (86 proteins). What was seen is that most (90%) of these 154 proteins (RLM+RLS, non-redundant) have negative GRAVY scores; hence they are more water soluble. This is in line with the assumption that recovery problems from protein fractionation workflows were caused by solubility issues. Separating complex mixtures of proteins is challenging, especially for samples containing membrane proteins such as liver microsomes/S9 fractions.

Using the 2D-SCX-RP method (combining trypsin + pepsin digestion results), 17566 unique peptides were detected (1% FDR) in RLM leading to 1120 identified proteins (1% FDR), with ~ 30% average protein sequence coverage, out of which 212 proteins (19%) were predicted to be transmembrane proteins (based on analysis using Phobius). From RLS samples, 15905 unique peptides and 1037 proteins were found with 28% protein sequence coverage. The S9 fractions also contained less transmembrane proteins (102, equivalent to 10%), which was anticipated, since both cytosolic and microsomal proteins are present (Duffus *et al.*, 2007). Both sample types yielded the same ranking of performance between 1D, 2D-CATWAX-RP and 3D workflows (Table 3.2). The 2D-SCX-RP approach provided the highest number of these unique proteins (486 in RLM, 440 in RLS). A total number of 170 in RLM and 96 in RLS were common to all four workflows (Figure 3.4). Moreover, if average sequence coverage is compared for these shared proteins, the SCX-RP method was again superior (61% vs 43% for 1D, 22% for CATWAX-RP, and 23% for 3D).

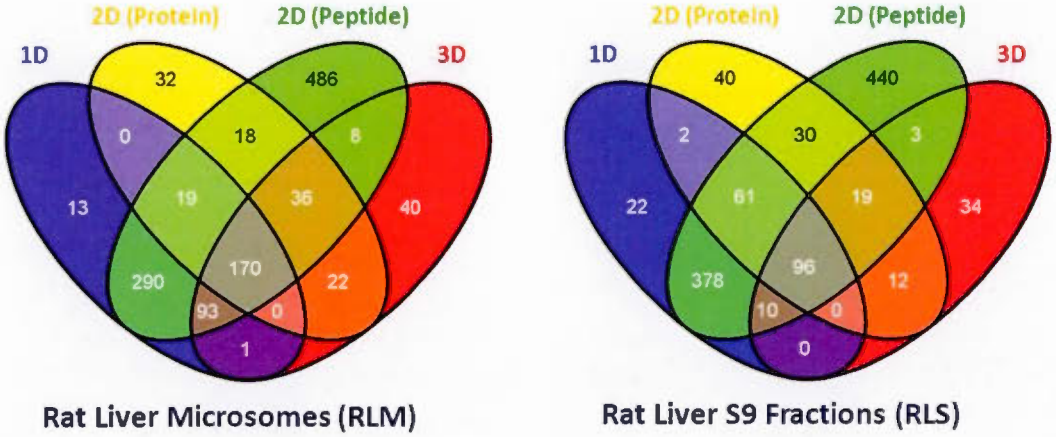


Figure 3.4 Venn diagrams representing the number of identified proteins in RLM (left) and RLS (right) using 1D-, 2D- and 3D-LC-MS/MS workflows

3.4.7 Rat Proteome Results

The rat liver proteome has been widely studied with the aim of improving existing knowledge on this specialized tissue. Several recent studies were performed to address questions in pathogenesis and development of liver cancer and chronic diseases (Low *et al.*, 2013; Hu *et al.*, 2013; Cao *et al.*, 2013), drug evaluation (Wang *et al.*, 2013) and drug-induced liver injury (van Swelm *et al.*, 2014; H. Zhang *et al.*, 2013; A. Zhang *et al.*, 2013; Sharanova *et al.*, 2013), and understanding of cellular functions (Bakala *et al.*, 2013; Chapel *et al.*, 2013; Gronemeyer *et al.*, 2013). Most recently, in an extensive study on whole rat liver homogenate, using an integrated "omics" approach on two rat strains (BN-Lx and SHR) combining multiple proteases (trypsin, Lys-C, Glu-C, Asp-N, and chymotrypsin), SCX fractionation and LC-MS/MS analysis on two platforms (TripleTOF 5600 and LTQ-Orbitrap Velos), Low *et al.* (2013) obtained peptide evidence for 26463 proteins with an overall sequence coverage of 15.6%. This impressive result was obtained based on searching a custom-built database with the goal of achieving a complete inventory of genetic variation,

combining data from genome and RNA sequencing, which includes genetic polymorphisms and post-transcriptional events. For comparative purposes, the analysis presented here was performed on Sprague-Dawley rat liver fractions with two proteases and one MS/MS platform, using the UniProt KB/Swiss-Prot database, comprising total of 7887 reviewed protein entries.

The current study incorporated a total of 491,615 (RLM) and 426,812 (RLS) high-resolution MS/MS spectra from the four strategies leading to the identification of 1185 (RLM) and 1081 (RLS) proteins at 1% FDR, of which 796 proteins (54%) were common between the two datasets, while 674 proteins (46%) were unique to either microsomes or S9 fraction. Combining the two datasets and removing redundancies, a total number of 1400 proteins were identified in the studied rat liver fractions, of which 1235 (88.2%) were also reported by Low *et al.* (2013).

Using Phobius (combined transmembrane topology and signal peptide predictor) (Krogh *et al.*, 2007), 215 and 103 IMPs was found in RLM (18% identified proteins) and RLS (10% identified proteins), respectively. An average of 5 and 7 peptides (confidence > 95%) were detected for each protein in RLM and RLS, respectively. However, the average protein sequence coverage was similar between RLM (31.4%) and RLS (29.9%) samples. Sequence coverage is of particular importance for studies involving protein modifications, since the higher the coverage, the better the chances of detecting covalent modifications on a target protein. Our group is particularly interested in elucidating novel protein targets of reactive molecules and therefore this was set as a high priority when optimizing the proteomic analysis workflow.

Characterizing the overall distribution of pI, molecular weights, and GRAVY scores is of interest when optimizing sample preparation steps (ion exchange, SPE, etc.) for increasing coverage of the proteomic analysis. Therefore, an in-depth analysis of the identified proteins was performed on the rat liver samples. From the

compiled dataset, RLM and RLS proteins were analyzed in terms of several physicochemical properties. A comparison of protein pI revealed that the majority of rat liver proteins have a theoretical pI in the range of 4–12 with 43.4% (RLM) and 34.8% (RLS) with basic characteristics (pI > 7). Considering the MW distribution, 2–3% of proteins were below 10 kDa, 60–62% in the range from 10 to 50 kDa, 29–30% in the range from 50 to 100 kDa, and 6–7% above 100 kDa, with an average molecular mass of 48.2 and 50.1 kDa for RLM and RLS, respectively. The identified proteins were also analyzed in terms of hydropathy (as measured by GRAVY scores). With a GRAVY score distribution in the range from -2.0 to 1.1, 16.8% of RLM and 12.9% of RLS proteins were characterized as hydrophobic (GRAVY > 0), coinciding well with the %IMPs predicted by Phobius (Figure 3.5). The identified proteins in rat liver samples were also analyzed in terms of biological properties using PANTHER (Thomas *et al.*, 2003). As depicted in Figure 3.6, GO analysis classified RLM and RLS proteins into 10 groups based on their molecular functions, over 45% of which represent catalytic functions.

Interestingly, there were more transporter and receptor proteins identified in the microsomes than in the S9 fractions, whereas the latter was richer in structural, molecule-binding, translation and enzyme regulator proteins. When classified by biological processes, 12 groups were identified with the highest proportion being involved in metabolic and cellular processes. However, a slightly higher proportion of the microsomal proteins were involved in localization and multicellular organismal processes, while more proteins were associated to biological regulation, cellular, developmental and metabolic processes from the S9 fractions. Many of these proteins play a crucial role in the metabolism of xenobiotics and thus are interesting for toxicology and drug discovery studies.

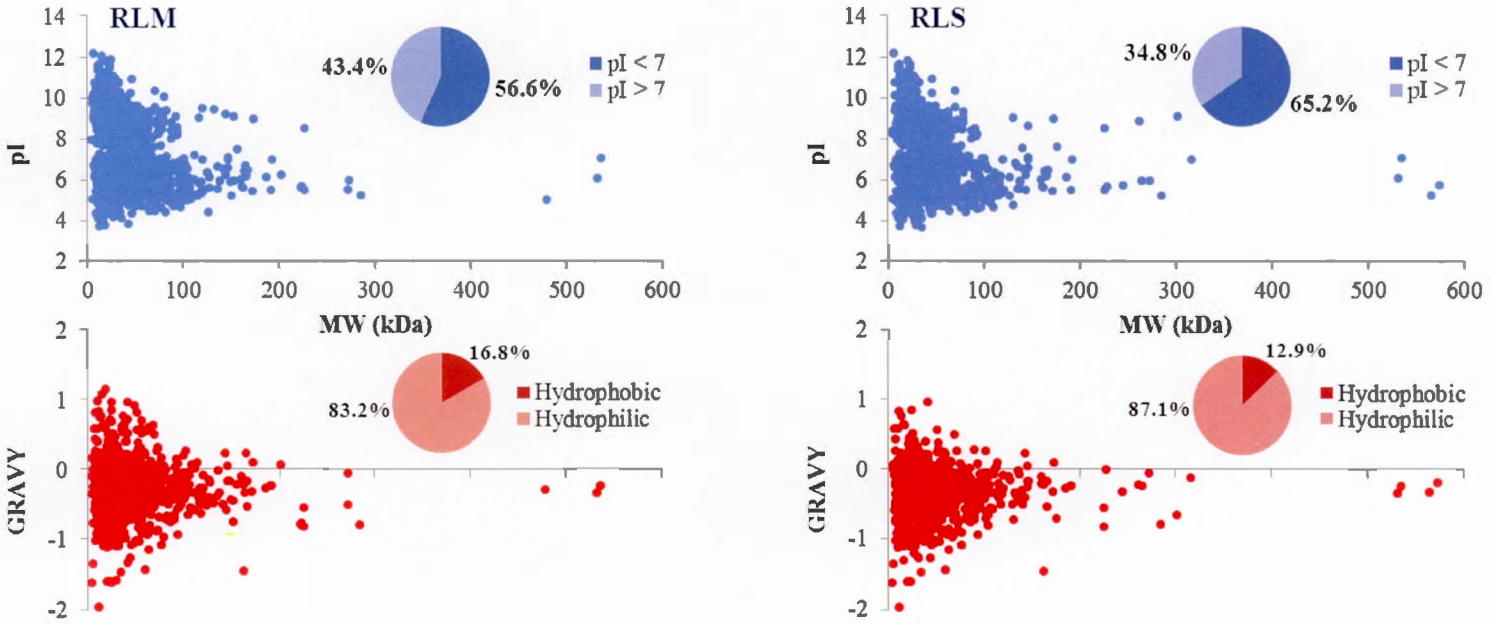


Figure 3.5 Statistical distribution of isoelectric point (above) and GRAVY score (below) as a function of MW for the proteins identified in RLM (right) and RLS (left)

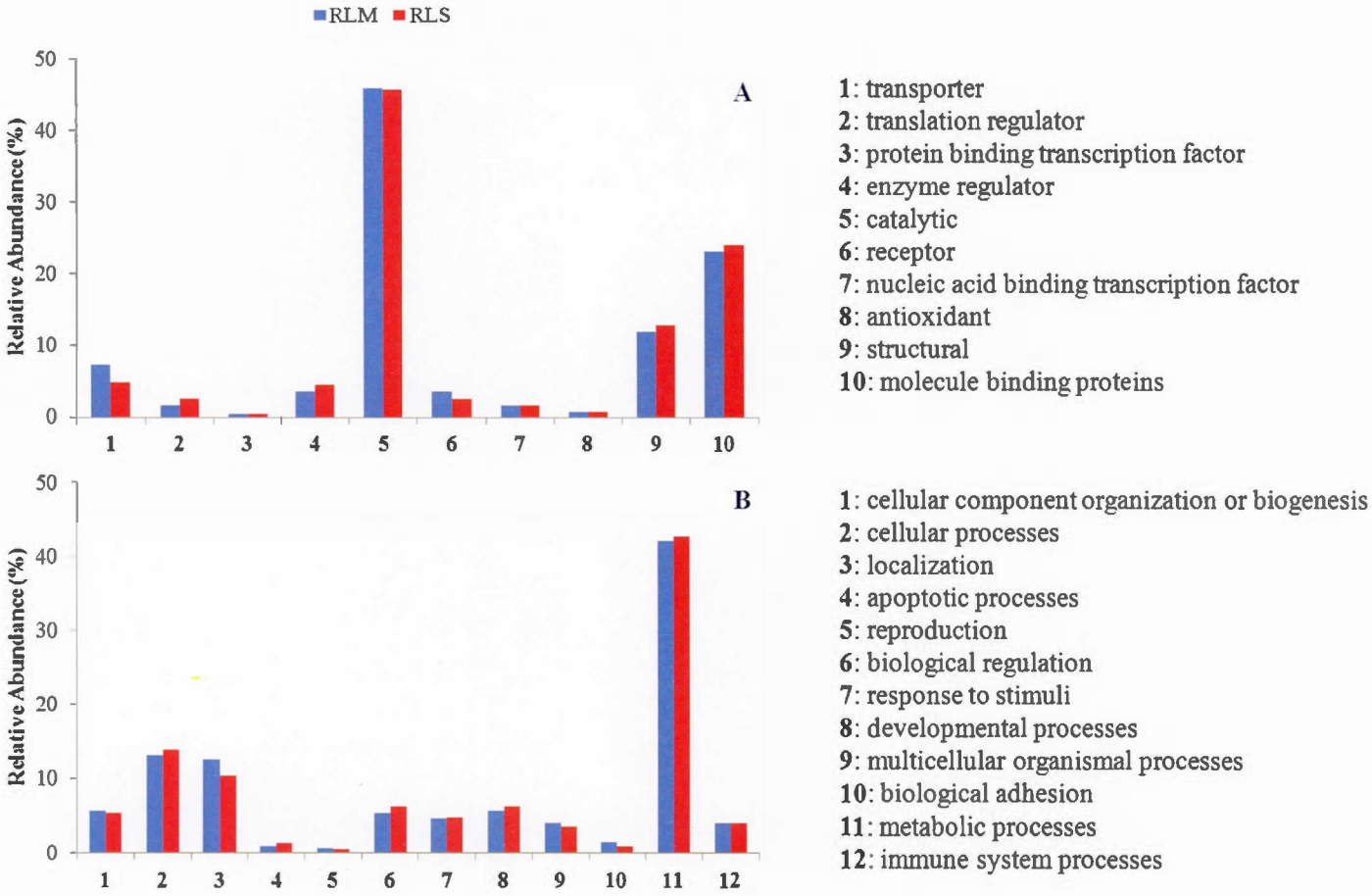


Figure 3.6 Gene ontology annotations based on molecular function (A) and biological process (B) for the proteins identified in RLM and RLS

3.4.8 Cross-Species Comparison of Liver Proteins

Rat and mouse are common animal models for various applications related to human metabolism, pharmacology and toxicology. Cross-species studies are informative for extrapolation of these models to humans (Benson and Giulio, 2006). Cross-species proteomics can also contribute to characterizing proteins of unknown function, especially between closely-related species which often have important sequence similarities (Wright *et al.*, 2010). Using the workflow optimized for the analysis of rat samples, mouse and human liver fractions were also analyzed and the datasets were compared. Table 3.5 summarizes the results from the analysis of microsomal and S9 fractions of all three species. Overall, 1400 distinct proteins were identified at 1% FDR in rat, 1791 in mouse, and 1939 in human with an average sequence coverage of 23-30%, an average of 8-10 peptides (confidence > 95%) per protein, and a reproducibility of 62-80% (for separately-fractionated duplicate samples). Also, similar to rat, microsomes contained a higher number of IMPs than S9 fractions in both mouse and human. As expected, the proteins identified in the three species demonstrated a high amount of similarity in terms of both physicochemical and biological properties, as illustrated in Figure 3.7 and Figure 3.8. As for isoelectric points, 38.6-43.1% of the proteins have pI > 7, whereas 15.6-16.4% could be classified as hydrophobic, coinciding well with the number of IMPs in each sample. Moreover, GO classification revealed that rat, mouse and human liver proteins are analogous in terms of function and involvement in biological processes, as expected.

Table 3.5 Proteomic analysis of rat, mouse and human liver microsomes and S9 fractions using the optimized 2D-LC-MS/MS method

Species	Sample	Proteins [a]	Peptides [a]	IMP (%)	Sequence % coverage [b]	Peptides per protein [b]	Total Proteins [a]
Rattus norvegicus	Microsomes	1120	17566	217 (19.4)	29.9	10	1400
	S9 fraction	1037	15905	104 (10.0)	28.4	10	
Mus musculus	Microsomes	1582	19262	294 (18.6)	24.9	8	1791
	S9 fraction	1538	18089	271 (17.6)	22.9	8	
Homo sapiens	Microsomes	1516	18739	308 (20.3)	24.0	8	1939
	S9 fraction	1570	20435	188 (12.0)	24.2	8	

[a] Number of proteins and peptides at 1% FDR

[b] Average of all identified proteins

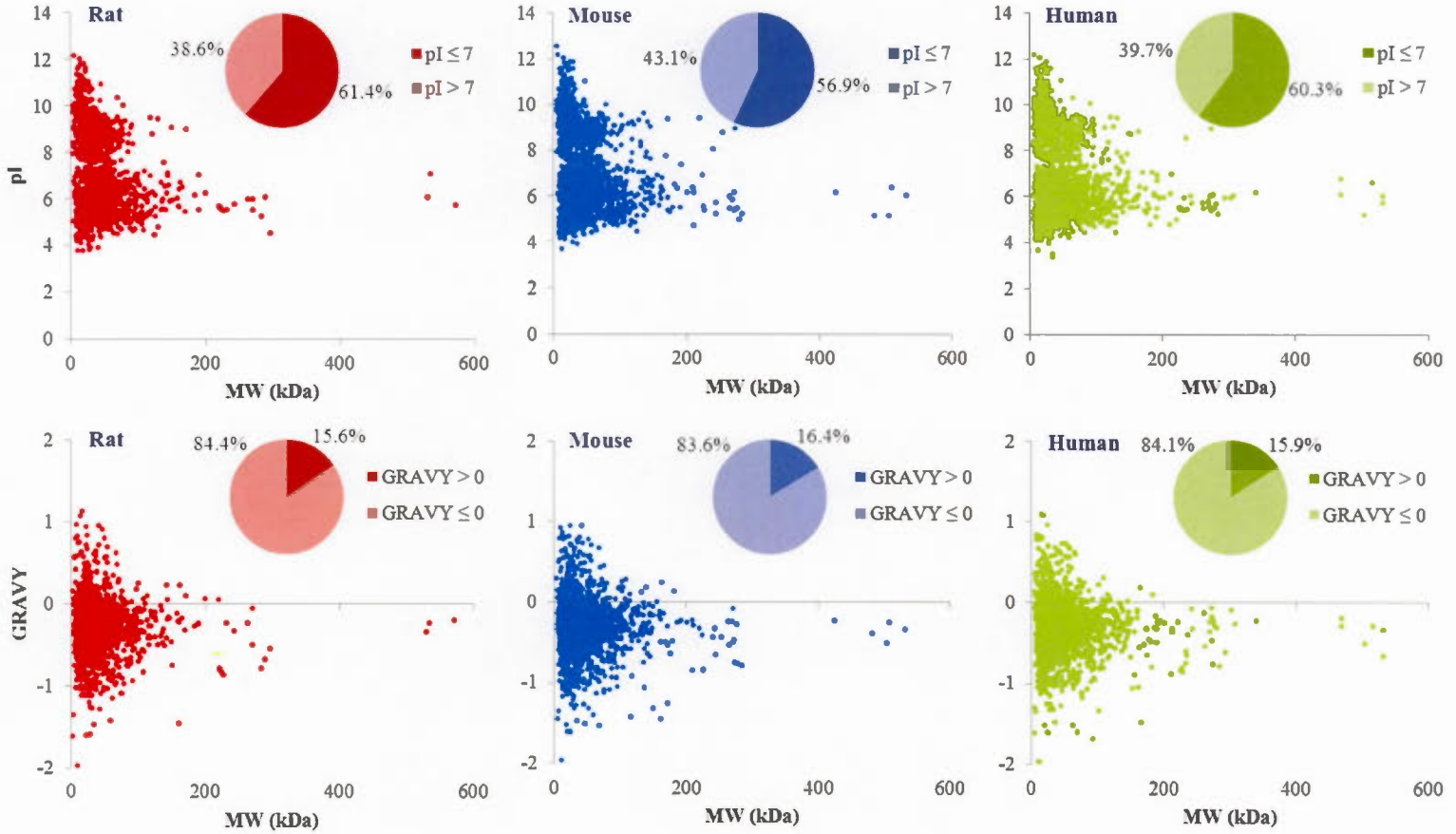


Figure 3.7 Statistical distribution of liver proteins in rat, mouse and human in terms of isoelectric point (above) and GRAVY score (below) as a function of MW

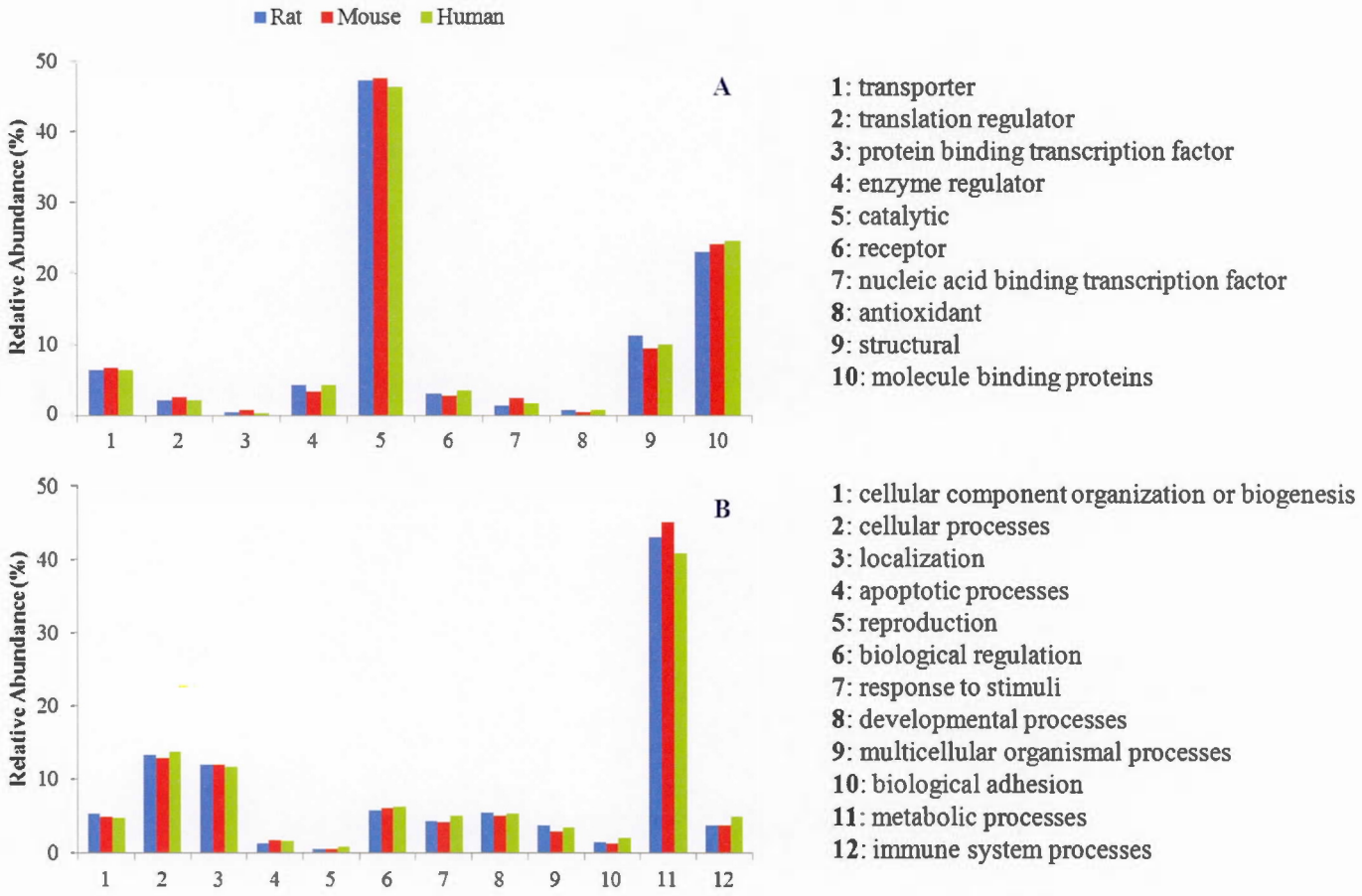


Figure 3.8 Gene ontology annotations based on molecular function (A) and biological process (B) for the liver proteins (from microsomes and S9) in rat, mouse and human

As shown in Figure 3.9, when a cross-species analysis was performed, comparing gene IDs only, on all datasets, 31.3% and 28.0% orthologous proteins were found between the three species for microsomal and S9 fractions, respectively, while approximately 44% were deemed as unique to one species. The combined results from microsomal and S9 fractions are shown in Figure 3.10. MSA analysis, performed to assess the non-orthologous proteins, demonstrated that for 22.6% (microsomal) and 24.6% (S9) of these proteins had reasonable sequence similarities, with at least 10% inter-species homology (Figure 3.11). Taking into consideration all the identified proteins (microsomal and S9), MSA analysis revealed that the set of human proteins are slightly more homologous to rat than mouse (Figure 3.12). Percent identity matrices (PIMs) and phylogenetic trees from this analysis are available in Supplemental Table S4.

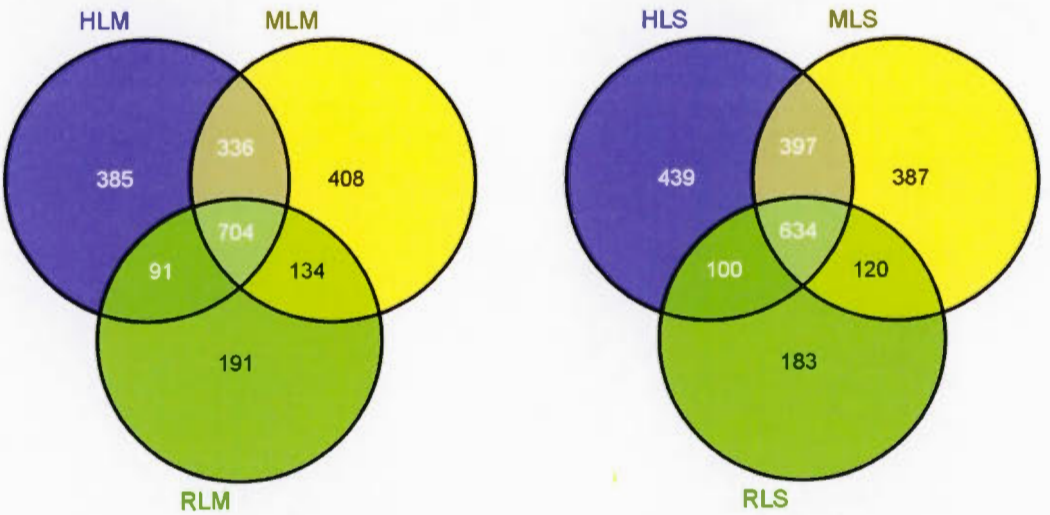


Figure 3.9 Venn diagrams depicting the number of shared and unique proteins identified in rat, mouse and human liver microsomal (LM) and liver S9 (LS) proteins

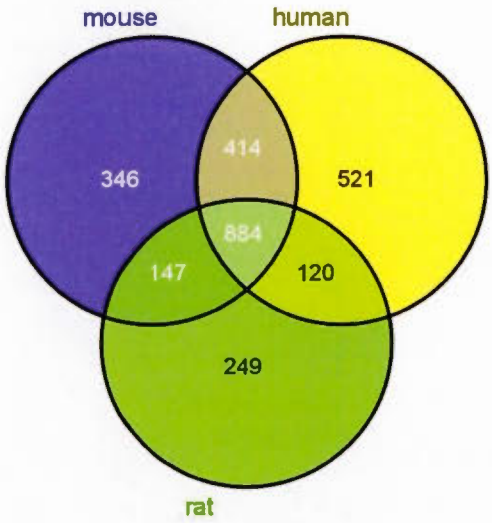


Figure 3.10 Venn diagrams depicting the number of shared and unique proteins identified in rat, mouse and human liver samples (microsomes and S9 fractions)

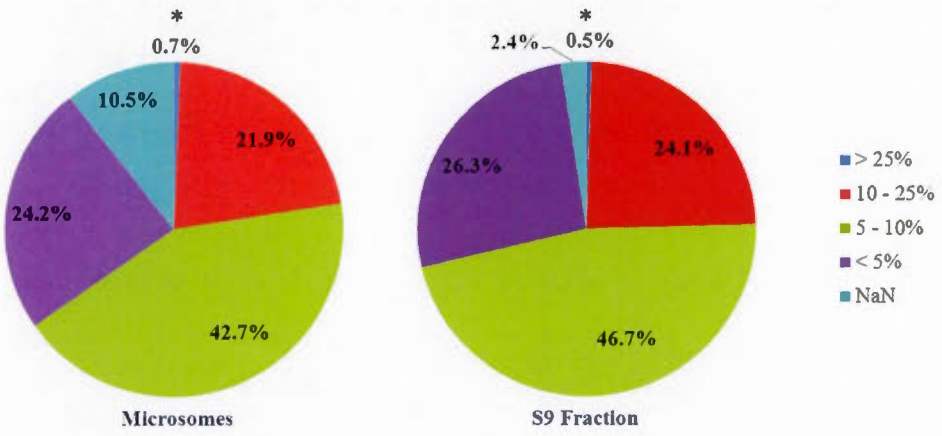


Figure 3.11 Percent alignment of proteins unique to rat, mouse or human liver microsomes (left) or S9 fraction (right). Labels (from the legend on right) are distributed in the pie charts from the * in a clockwise manner (The term NaN stands for "not a number")

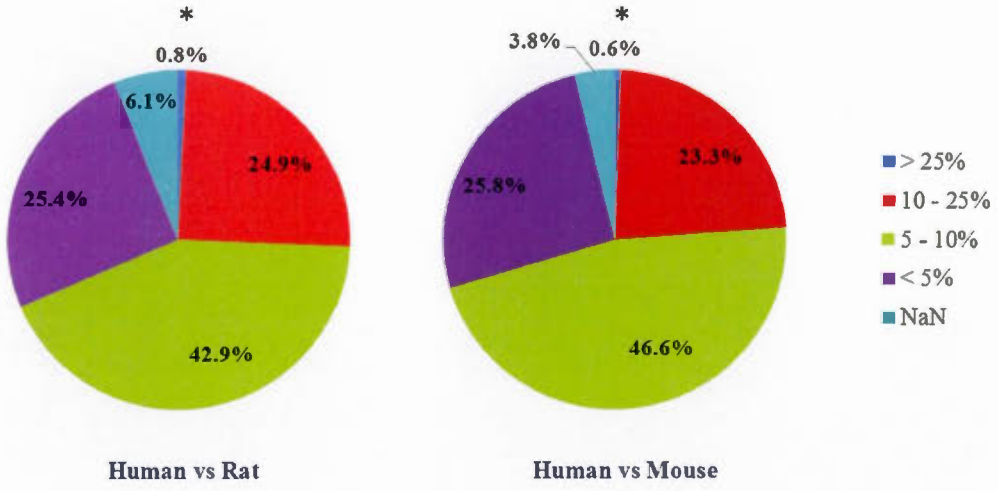


Figure 3.12 Percent alignment of proteins unique to human with those unique to rat (left) and mouse (right). Labels (from the legend on right) are distributed in the pie charts from the * in a clockwise manner (The term NaN stands for "not a number")

A complete list of the proteins identified (at 1% FDR) in each sample, and the distinct peptides detected for each protein is published as a Data in Brief article (Golizeh *et al.*, 2015b). In this related article, links to publicly accessible repositories are provided. In an effort to quantify the impact of making this data publicly available, the lists of peptides from tryptic digestion of human, mouse and rat samples were compared to the most recent database entries on PeptideAtlas (2013-11, 2014-12 and 2014-08 builds for rat, mouse and human, respectively) (Farrah *et al.*, 2013). Remarkably, a relatively large proportion of our highly-confident peptides were not found in the most current version of PeptideAtlas (20.1% of the human peptides, and 35.8% for each of the mouse and rat peptide lists). Therefore, once incorporated, this large dataset would add a significant amount of additional information into this publicly accessible database. These data would be useful for subsequent studies involving quantitative proteomics applications as well as further probing of biological information contained in these rich datasets.

Liver microsomes and S9 fractions are routinely used in drug metabolism studies and therefore, an in-depth characterization of the proteins contained in these fractions is of particular interest for probing protein modifications resulting via drug bioactivation (reactive metabolites). Several studies have been performed with the goal of identifying protein targets of reactive compounds in liver microsomes (Yang *et al.*, 2014; Tzouros and Pähler, 2009; Rombach and Hanzlik, 1999; Shin *et al.*, 2007), and we anticipate this comprehensive proteomic analysis of these three species will facilitate future work in this challenging area.

3.5 Conclusions

Sample preparation and enrichment is a crucial step in the analysis of complex samples and directly affects the efficiency of any proteomic workflow. In this study, a parallel dual digestion method followed by a comparison of different fractionation strategies prior to UHPLC-MS/MS was employed in the analysis of rat liver microsomal and S9 fractions. Protein-level fractionation using mixed-bed weak cation/anion exchange (CATWAX) did not improve proteome coverage for these samples, compared to the 1D(RP)-LC-MS/MS workflow and did not ameliorate results when combined with peptide-level SCX fractionation either. The 2D(SCX-RP)-LC-MS/MS approach yielded the best results and was therefore chosen for the comparative analysis of rat, mouse and human liver fractions. Informative results including physicochemical properties and GO classification of the identified proteins, as well as an overall comparison of homology between the three species was performed. This study resulted in large sets of high-resolution MS/MS data yielding impressive proteome coverage for challenging liver fractions containing a relatively high number of membrane proteins. The results of this work are of particular interest to the drug discovery and toxicology communities. This information could be

specifically useful in several applications, for example cross-species extrapolation of toxicity, characterizing protein modifications, and developing targeted quantitative assays for verification of potential biomarker panels.

3.6 Acknowledgements

M. G. would like to thank GRASP (*Groupe de Recherche Axé sur la Structure des Protéines*, FRSQ) and *la Fondation de l'UQÀM* (*Bourses d'Excellence des Cycles Supérieurs* FARE) scholarships. Authors would also like to thank Dr. Sean Seymour (AB Sciex), Dr. Eric Deutsch (Institute for System Biology / PeptideAtlas) and the PRIDE team for their kind and timely support.

3.7 Supporting Information

Multiple sequence alignment (MSA) analysis results of the proteins unique to rat, mouse or human (available online).

CHAPTER FOUR

DATASET FROM PROTEOMIC ANALYSIS OF RAT, MOUSE, AND HUMAN LIVER MICROSOMES AND S9 FRACTIONS

Makan Golizeh, Christina Schneider, Leanne B. Ohlund and Lekha Sleno

Published in **Data in Brief 2015, 3, 95-98**. Supplementary data available online at doi:10.1016/j.dib.2015.02.007.

Large amount of proteomics data are available via public data repositories. These data provide insight into mechanisms involved in biological processes and facilitate validation of findings across different areas of research. The analytical workflow developed in Chapter 3 enabled efficient proteomics analysis of liver S9 and microsomal fractions in rat, mouse and human. Raw and processed data from this analysis were deposited onto two MS-based proteomics public data repositories. These data can be used in cross-species comparison between proteins found in the three species, extrapolation of toxicity data from rat and mouse to human, or more targeted analysis, such as MRM-based assays.

Makan Golizeh, Christina Schneider, Leanne Ohlund and Lekha Sleno are co-authors of this article. Makan Golizeh conducted the experiments, processed and submitted the data, and prepared the original manuscript. Christina Schneider and Leanne Ohlund participated in the analytical method development. Professor Lekha Sleno supported and supervised the project, verified data processing, revised and finalized the manuscript.

4.1 Abstract

Rat, mouse and human liver microsomes and S9 fractions were analyzed using an optimized method combining ion exchange fractionation of digested peptides, and ultra-high performance liquid chromatography (UHPLC) coupled to high resolution tandem mass spectrometry (HR-MS/MS). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaíno *et al.*, 2013) with the dataset identifiers **PXD000717**, **PXD000720**, **PXD000721**, **PXD000731**, **PXD000733** and **PXD000734**. Data related to the peptides (trypsin digests only) were also uploaded to Peptide Atlas (Farrah *et al.*, 2013) and are available with the dataset identifiers **PASS00407**, **PASS00409**, **PASS00411**, **PASS00412**, **PASS00413** and **PASS00414**. The present dataset is associated with a research article published in EuPA Open Proteomics (Golizeh *et al.*, 2015c).

4.2 Value of the Data

Liver proteins were identified in microsomal and S9 fractions with high sequence coverage. Comprehensive list of proteotypic peptides was reported for each identified protein enables more targeted analyses of these proteins. Data acquired from cross-species analysis of rat, mouse and human liver microsomes and S9 fractions. Multiple-sequence alignment (MSA) analysis can be done between proteins exclusively found in each species.

4.3 Specifications Table

Subject area	Chemistry, biology
More specific subject area	Proteomic analysis
Type of data	1) Raw and processed mass spectrometry data acquired by 2D-LC-MS/MS analysis of rat, mouse, and human liver microsomes and S9 fractions 2) Excel datasheets with identified proteins and corresponding peptides from each analyzed sample
How data was acquired	2D-LC-MS/MS using Agilent 1200 HPLC, Shimadzu Nexera UHPLC, and AB Sciex TripleTOF 5600 mass spectrometer
Data format	*.wiff*, *.wiff.scan* (raw files) *.group*, *.xml* (search files) *.mgf* (peak files) *.xlsx* (FDR analysis and processed results)
Experimental factors	No sample pre-treatment applied
Experimental features	Liver microsomes/S9 fractions were solubilized, denatured, and subjected to a trypsin/pepsin parallel dual-digestion. Digested peptides were fractionated using strong cation exchange (SCX) chromatography prior to UHPLC-MS/MS with subsequent data-mining and bioinformatics analysis.
Data source location	Université du Québec à Montréal (UQÀM), Chemistry Department, Montréal, QC, Canada

Data accessibility	The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Vizcaíno *et al.*, 2013) with the dataset identifiers **PXD000717**, **PXD000720**, **PXD000721**, **PXD000731**, **PXD000733** and **PXD000734**. Data related to the peptides (trypsin digests only) were also uploaded to Peptide Atlas (Farrah *et al.*, 2013) and are available with the dataset identifiers **PASS00407**, **PASS00409**, **PASS00411**, **PASS00412**, **PASS00413** and **PASS00414**.

4.4 Experimental Design, Materials and Methods

Rat, mouse, or human liver microsomes or S9 fractions (0.6 mg protein, $n = 2$) were solubilized in 2% SDS solution (1:1 v/v ratio) and then diluted with 0.1 M ammonium bicarbonate (pH 8.5) prior to reductive alkylation with dithiothreitol (2.5 mM) and iodoacetamide (5 mM). Additional ammonium bicarbonate (for trypsin), or 0.2% trifluoroacetic acid in 20% methanol (for pepsin), was added for an overnight digestion at a 1:50 (w/w) enzyme/protein ratio. Digests were neutralized, diluted with water, and subjected to solid-phase extraction (SPE) on a 1 cc (30 mg) OASIS HLB cartridge (Waters, Milford, MA), eluting with 100% methanol (1 ml). Eluates were evaporated to dryness under vacuum, reconstituted in SCX buffer A (see below), and injected (100 µl, 0.5 mg protein) onto a Zorbax 300-SCX 150 × 2.1 mm column with 5 µm (300 Å) particles (Agilent Technologies, Palo Alto, CA) using an Agilent 1200 series HPLC equipped with a binary pump, degasser, diode array detector and fraction collector. SCX fractionation was performed (250 µl/min) with a

gradient of 0–50% B in 15 min, up to 100% B at 25 min, then held for an additional 5 min at 100% B, where buffers A and B were 10 mM potassium dihydrogen phosphate in 25% acetonitrile (pH 2.75), and 1 M potassium chloride in buffer A (pH 2.75), respectively. UV absorbance was monitored at 220 and 280 nm. For trypsin samples, 3 min (0.75 ml) fractions were aliquoted into 1.5 ml tubes between 1.5 and 19.5 min, while for pepsin, 4 min (1.0 ml) fractions were collected between 1.5 and 25.5 min. Fractions were evaporated to dryness under vacuum and kept at −30°C.

Dried fractions were reconstituted in 10% acetonitrile (100 µl) and injected (20 µl) onto an Aeris PEPTIDE XB-C_{18} 100 × 2.1 mm column, with solid core 1.7 µm particles (100 Å) (Phenomenex, Torrance, CA). RP-LC was performed (300 µl/min, 40°C) on a Nexera UHPLC system (Shimadzu, Columbia, MD) with water (A) and acetonitrile (B), both containing 0.1% formic acid with a gradient of 5% B held for 2 min, increased linearly to reach 30% B at 24 min, to 50% B at 26 min, then to 85% B at 26.5 min and held for 2 min. MS and MS/MS spectra were collected on a high-resolution hybrid quadrupole-time-of-flight (QqTOF) TripleTOF 5600 mass spectrometer (AB Sciex, Concord, ON) equipped with a DuoSpray ion source in positive ion mode. The instrument performed a survey TOF-MS acquisition from *m/z* 140–1250 (250 ms accumulation time), followed by MS/MS on the 15 most intense precursor ions from *m/z* 250–1250 (precursors excluded for 20 seconds after two occurrences) using information-dependent acquisition (IDA) with dynamic background subtraction (DBS). Each MS/MS acquisition (*m/z* 80–1500) had an accumulation time of 50 ms and collision energy of 30 ± 10 V. The total cycle time was 1.05 s.

MS/MS files were combined and searched against the UniProt protein database (http://www.uniprot.org, release date 26/06/2013) by ProteinPilot software (version 4.1) using Paragon algorithm (Shilov *et al.*, 2007) using a thorough ID search with no specified enzyme and carbamoylation as a fixed cysteine modification.

The search was performed for +2 to +4 charge states and MS tolerance was 0.05 Da on precursor ions and 0.1 Da on fragment ions. All duplicates were first processed alone, then together and finally tryptic and peptic digest for each sample were co-processed to obtain the total number of proteins and peptides. Proteins were identified with a 1% global false discovery rate (FDR) using a target-decoy database search algorithm (Tang *et al.*, 2008) in ProteinPilot Descriptive Statistics Template (version 3.001p) (http://www.absciex.com/PDST). ProteinPilot Protein Alignment Template (version 2.000p) was also used for replicate analysis. The list of UniProt accession numbers from identified proteins was uploaded to NCBI Batch Entrez (http://www.ncbi.nlm.nih.gov/sites/batchentrez) to obtain the batch FASTA file, which was subsequently submitted to ExPASy (http://www.expasy.org) for determination of isoelectric point and monoisotopic molecular weight, to Phobius (http://phobius.sbc.su.se) (Krogh *et al.*, 2007) for prediction of integral membrane proteins, and to GRAVY Calculator (http://www.gravy-calculator.de) to compute grand average of hydropathy (GRAVY) scores. Lists of the identified proteins and the corresponding proteotypic peptides in rat, mouse, and human liver microsomes and S9 fractions can be found in Supplemental Tables S1 and S2 (see Table 4.1 for details). The related MS data are also available via public proteomics data repositories (see Table 4.2 for the accession numbers).

4.5 Acknowledgements

Authors would like to thank Dr. Sean Seymour (AB Sciex), Dr. Eric Deutsch (Institute for System Biology / PeptideAtlas) and the PRIDE team for their kind and timely support.

4.6 Supporting Information

Table S1, final list of the proteins identified in rat, mouse and human liver microsomes/S9 fractions at 1% FDR; Table S2, final list of the prototypic peptides identified in rat, mouse and human liver microsomes/S9 fractions at 1% FDR.

Table 4.1 Processed data from proteomic analysis of rat, mouse, and human liver microsomes and S9 fractions

Spices	Sample	ID	Protein Data Location	Peptide Data Location
Rattus norvegicus	Liver microsomes	RLM	Supplemental Table S1-1	Supplemental Table S2-1
	Liver S9 fraction	RLS	Supplemental Table S1-2	Supplemental Table S2-2
Mus musculus	Liver microsomes	MLM	Supplemental Table S1-3	Supplemental Table S2-3
	Liver S9 fraction	MLS	Supplemental Table S1-4	Supplemental Table S2-4
Homo sapiens	Liver microsomes	HLM	Supplemental Table S1-5	Supplemental Table S2-5
	Liver S9 fraction	HLS	Supplemental Table S1-6	Supplemental Table S2-6

Table 4.2 Dataset identifiers of the mass spectrometry data obtained from the analysis of rat, mouse, and human liver microsomes and S9 fractions on the public proteomics repositories

Spices	Sample	ID	Protein Dataset Identifier (PRIDE)	Peptide Dataset Identifier (PeptideAtlas)
Rattus norvegicus	Liver microsomes	RLM	PXD000720	PASS00407
	Liver S9 fraction	RLS	PXD000717	PASS00409
Mus musculus	Liver microsomes	MLM	PXD000731	PASS00411
	Liver S9 fraction	MLS	PXD000733	PASS00412
Homo sapiens	Liver microsomes	HLM	PXD000721	PASS00413
	Liver S9 fraction	HLS	PXD000734	PASS00414

CHAPTER FIVE

IDENTIFICATION OF ACETAMINOPHEN ADDUCTS OF RAT LIVER MICROSOMAL PROTEINS USING 2D-LC-MS/MS

Makan Golizeh, André LeBlanc and Lekha Sleno

Published in **Chemical Research in Toxicology 2015, 28, 2142-2150**. Supporting information included in Appendix A and Appendix B (also available online at doi:10.1021/acs.chemrestox.5b00317).

The ultimate goal of this dissertation was to develop an analytical method for the identification of reactive metabolite protein targets using LC-MS/MS. In this chapter, rat liver microsomes were treated with acetaminophen, a well-studied model in liver proteins adduction research, and then subjected to a 2D-LC-MS/MS workflow. An extensive data processing approach was later established to detect potential covalent modifications, efficiently remove false positives, and characterize the adducted peptides. The results presented in this chapter can potentially improve the existing knowledge of the mechanisms governing acetaminophen-induced liver injury, and the developed method can be used in similar studies involving protein adduction and drug-related toxicity.

Makan Golizeh, André LeBlanc and Lekha Sleno are co-authors of this article. Makan Golizeh conducted the literature survey, prepared experimental protocols, conducted the experiments, processed the data, and prepared the original manuscript. André LeBlanc developed the microsomal incubation procedure and participated in interpretation of the results. Professor Lekha Sleno supported and supervised the project, verified data analysis and interpretation of the results, revised and finalized the manuscript.

5.1 Abstract

Xenobiotic metabolism in the liver can give rise to reactive metabolites that covalently bind to proteins, and determining which proteins are targeted is important in drug discovery and molecular toxicology. However, there are difficulties in the analysis of these modified proteins in complex biological matrices due to their low abundance. In this study, an analytical approach was developed to systematically identify target proteins of acetaminophen (APAP) in rat liver microsomes (RLM) using two-dimensional chromatography and high-resolution tandem mass spectrometry. *In vitro* microsomal incubations, with and without APAP, were digested and subjected to strong cation exchange (SCX) fractionation prior to reversed-phase UHPLC-MS/MS. Four data processing strategies were combined into an efficient label-free workflow meant to eliminate potential false positives, using peptide spectral matching, statistical differential analysis, product ion screening, and a custom-built delta-mass filtering tool to pinpoint potential modified peptides. This study revealed four proteins, involved in important cellular processes, to be covalently modified by APAP. Data are available via ProteomeXchange with identifier **PXD002590**.

5.2 Introduction

Drug-induced liver injury (DILI) is a major obstacle in drug development which often leads to drug use restriction, post-market withdrawal, and attrition due to toxicity (Kaplowitz, 2005). Approximately 1000 drugs, herbal products, dietary supplements and illicit compounds are known to be associated with liver injury (Chen *et al.*, 2014). However, although intrinsic hepatotoxicity is reasonably predictable from pharmacology and chemical structures, the pathophysiological mechanisms of DILI are poorly understood (Fontana, 2014). Several therapeutic drugs with different

structures and mechanisms of action have been used in preclinical tests to identify the risk of liver injury in humans (Chen *et al.*, 2014). Acetaminophen (APAP) is a widely used analgesic and antipyretic drug which is safe and effective at therapeutic doses but can act as a potent hepatotoxin at high doses, and thus is commonly used as a model to study DILI. The APAP model has several advantages, including being dose-dependent and reproducible in animal models for clinically-relevant studies (Jaeschke *et al.*, 2011). Moreover, extensive research over the past 40 years has provided substantial information on the mechanisms of APAP-mediated hepatotoxicity (Abdelmegeed *et al.*, 2013; Jaeschke *et al.*, 2003; Jaeschke *et al.*, 2012; Jiang *et al.*, 2015; Xie *et al.*, 2015).

One of the liver's main physiological roles is the metabolism of xenobiotics into hydrophilic metabolites to facilitate their excretion from the body. The liver is therefore exposed to high concentrations of drugs and thus is a primary target for chemical-induced toxicity. At therapeutic doses, APAP is detoxified via glucuronylation and sulfation. However, a proportion of the drug undergoes bioactivation to *N*-acetyl-*p*-benzoquinoneimine (NAPQI), a reactive electrophile, which rapidly reacts with free thiol groups, and if not completely quenched by hepatic GSH, can covalently bind to protein thiols (Figure 5.1) (Yuan and Kaplowitz, 2013). Although protein covalent binding does not always lead to toxicity, modification of certain proteins by reactive drug metabolites can potentially alter critical cellular pathways causing or contributing to the extent and direction of toxicity (Liebler, 2008). Therefore, identification of the proteins targeted by reactive metabolites can help better understand the mechanism of drug-induced toxicity. However, since protein adducts are typically much less abundant compared to non-adducted proteins, their detection and identification in complex mixtures often pose serious analytical challenges (Tzouros and Pähler, 2009).

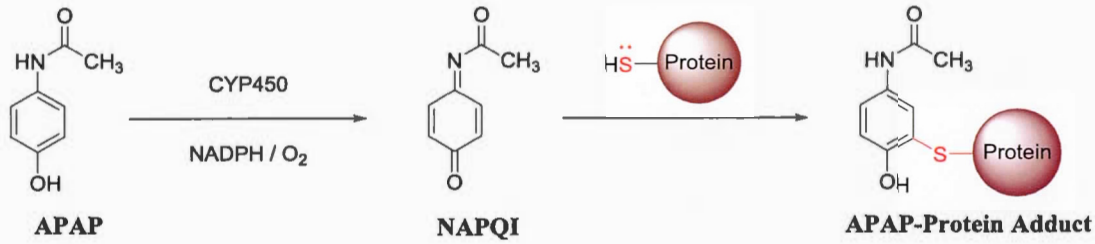


Figure 5.1 Covalent binding of NAPQI to the free thiol group of cysteine residues in proteins. NAPQI forms via phase I biotransformation of APAP by cytochrome P450 enzymes through NADPH-dependant oxidation.

In 1973, Jollow *et al.* showed that APAP-protein adducts accumulate in hepatocytes causing extensive necrosis after acetaminophen administration. Previous attempts to separate and identify APAP-modified proteins have employed immunohistochemical (Bartolone *et al.*, 1989; Roberts *et al.*, 1991), electrochemical (Muldrew *et al.*, 2002), or radiochemical (Qiu *et al.*, 1998) methods, which although they provided valuable information on the possible targets and mechanisms, required intensive labour, were unsuitable for the analysis of complex samples, and yielded no or little insight on the position of the modifications (Zhou, 2003; Park *et al.*, 2005).

The development of MS-based proteomics and multidimensional chromatographic separation in the late 1990s and early 2000s (Aebersold and Mann, 2003; Washburn *et al.*, 2001) has enabled high-throughput analysis of modified proteins with higher sensitivity and better selectivity. Mass spectrometry is currently the method of choice for identifying proteins and studying their post-translational modifications (Altelaar and Heck, 2012). The emergence of high-resolution MS (HR-MS) (Mehmood *et al.*, 2015), modern data acquisition techniques (Bern *et al.*, 2010; Gillet *et al.*, 2012), and multidimensional (Di Palma *et al.*, 2012) liquid chromatography, in conjunction with novel sample preparation (Feist and Hummon, 2015) and data analysis (Titz *et al.*, 2014) approaches have allowed significant enhancements to MS-based proteomics for the detection of low-abundance and less

soluble proteins and their post-translational modifications. In the present study, an analytical approach was developed combining solid-phase extraction (SPE) with strong cation exchange (SCX) and reversed-phase (RP) LC for sample preparation and an extensive data analysis workflow to identify the covalent binding target proteins of APAP in RLM. Four proteins were confirmed to be APAP-modified based on their accurate mass and high-resolution MS/MS peptide spectra.

5.3 Experimental

5.3.1 Chemicals and Reagents

Aroclor-induced male Sprague–Dawley RLM were purchased from Bioreclamation IVT (Baltimore, MD). Trypsin (TPCK-treated, from bovine pancreas), APAP, *n*-dodecyl-β-D-maltoside (DDM) and all other chemical reagents were purchased from Sigma-Aldrich (St. Louis, MO). Synthetic peptides were purchased from Biomatik (Cambridge, ON). HPLC-grade acetonitrile (ACN), and methanol were from Caledon (Georgetown, ON), and ultra-pure water was from a Millipore Synergy UV system (Billerica, MA).

5.3.2 Drug Metabolism

RLM (0.5 mg protein) were incubated (3 h, 37°C) ± APAP (0.1 mM), in the presence of NADPH (2 mM), in 100 mM phosphate buffer pH 7.4 (total volume 0.4 ml). GSH (0.12 mM) was then added to remove any excess reactive metabolite. Samples were prepared in triplicate.

5.3.3 Protein Digestion

DDM (0.1%, 15 min sonication, 25°C) was added to help solubilize microsomal proteins and reductive alkylation was performed using dithiothreitol (DTT, 5 mM, 20 min, 37°C) and iodoacetamide (IAM, 15 mM, 30 min, 37°C, dark), after adding 250 µl of 100 mM ammonium bicarbonate (pH 8.5). Samples were then diluted with ammonium bicarbonate buffer to a total volume of 0.75 ml and incubated with trypsin overnight (18 h, 37°C) at a 1:25 (w/w) enzyme-to-protein ratio. Digests were then cleaned-up by solid-phase extraction on 1 cc (30 mg) OASIS HLB cartridges (Waters, Milford, MA) with 100% methanol elution (1 ml). Eluate was dried under vacuum and stored at −30°C.

5.3.4 SCX Chromatography

Dried SPE extracts were reconstituted in buffer A (120 µl, see below) for injection (100 µl) onto a Zorbax 300-SCX 150 × 2.1 mm column with 5 µm (300 Å) particles (Agilent Technologies, Palo Alto, CA) using an Agilent 1200 series HPLC equipped with a temperature-controlled autosampler, binary pump, degasser, diode array detector and refrigerated fraction collector. SCX fractionation was performed at a flow rate of 250 µl/min with a gradient of 0–8% B in 6 min, increasing to 30% at 11 min, 65% at 20 min, 95% at 25 min, and held for 10 min at 95% B, where buffers A and B were 10 mM KH_2PO_4 in 25% ACN (pH 2.75), and 1 M KCl in buffer A (pH 2.75), respectively. UV absorbance was monitored at 220 and 280 nm. Eighteen (1.25 min) fractions were aliquoted from 1–23.5 min. Fractions were dried and reconstituted in 100 µl 10% ACN prior to LC-MS/MS analysis.

5.3.5 Microsomal Trapping Experiments with Standard Peptides

Glutathione, QACLFK, ILISDFGLCK, LQQCPFEDHVKL, MPFTSSCLII, FIDLIPTNLPHAVTCDIK, GKEGAHAPCASE and VFANPEDCAGFGK peptides were incubated (0.125 mM, 90 min, 37°C) with APAP (0.02 mM), NADPH (1 mM), and RLM (1.25 mg/ml protein) in 100 mM phosphate buffer pH 7.4 (total volume 0.4 ml). The reaction was quenched with 1:1 (v/v) ice-cold ACN, centrifuged (10 min, 14,000 g, 4°C), and supernatant was evaporated under vacuum and reconstituted in 200 µl 10% ACN prior to LC-MS/MS analysis.

5.3.6 RP-UHPLC-MS/MS Analysis

Samples were injected (20 µl) onto an Aeris PEPTIDE XB-C_{18} 100 × 2.1 mm column, with solid core 1.7 µm particles (100 Å) (Phenomenex, Torrance, CA) using a Nexera UHPLC system (Shimadzu, Columbia, MD) with water (A) and ACN (B), both containing 0.1% formic acid, at a flow rate of 300 µl/min (40°C). The gradient started at 5% B, held for 2.5 min, and was linearly increased to 30% B at 40 min, to 50% B at 50 min, then to 85% B at 55 min. MS and MS/MS spectra were collected on a high-resolution quadrupole-time-of-flight TripleTOF 5600 mass spectrometer (AB Sciex, Concord, ON) equipped with a DuoSpray ion source in positive mode set at 5 kV source voltage, 500°C source temperature and 50 psi GS1/GS2 gas flows, with declustering potential of 80 V. The instrument performed a survey TOF-MS acquisition from *m/z* 140–1250 (250 ms accumulation time), followed by MS/MS on the 15 most intense precursor ions from *m/z* 250–1250 (excluded for 20 seconds after two occurrences) using information-dependent acquisition (IDA) with dynamic background subtraction. Each MS/MS acquisition had an accumulation time of 50 ms and collision energy of 30 ± 10 V. The total cycle time was 1.05 s.

An IDA inclusion list was later generated in order to obtain MS/MS data on potential hits from non-spectral matching approaches. The method included a survey TOF-MS acquisition from *m/z* 250–1250, followed by MS/MS for the top ten precursor ions, with a total cycle time of 1.25 s. A representative UV trace (220 nm) from the SCX fractionation and the total ion chromatogram for a chosen fraction are displayed in Figure 5.2.

5.3.7 Data Analysis

LC-MS/MS files from each sample (18 SCX fractions combined) were searched against the UniProtKB/Swiss-Prot protein database (release date 07/01/2015) by ProteinPilot software for *rattus norvegicus* using the Paragon algorithm (Shilov *et al.*, 2007). In order to find potential APAP adducts, the protein search algorithm was changed to consider a probability of 0.85 for APAP modification on cysteine residues. The search was performed for +2 to +4 charge states at a MS tolerance of 0.05 Da on precursor ions and 0.1 Da on fragments. Proteins were identified with a 1% global false discovery rate (FDR) using a target-decoy database search algorithm (Tang *et al.*, 2008). MarkerView software was used to generate peak lists and to find features ($m/z > 300$) showing a fold-change > 3 in the APAP-treated sample compared to the control (−APAP), combining all fractions for each sample. MetabolitePilot software was employed to find LC-MS peaks with a sample-to-control signal ratio > 20 (one fraction at a time) including at least one of the four APAP-cysteine adduct-related diagnostic product ions (*m/z* 166.0321, 182.0270, 208.0427 and 225.0692).

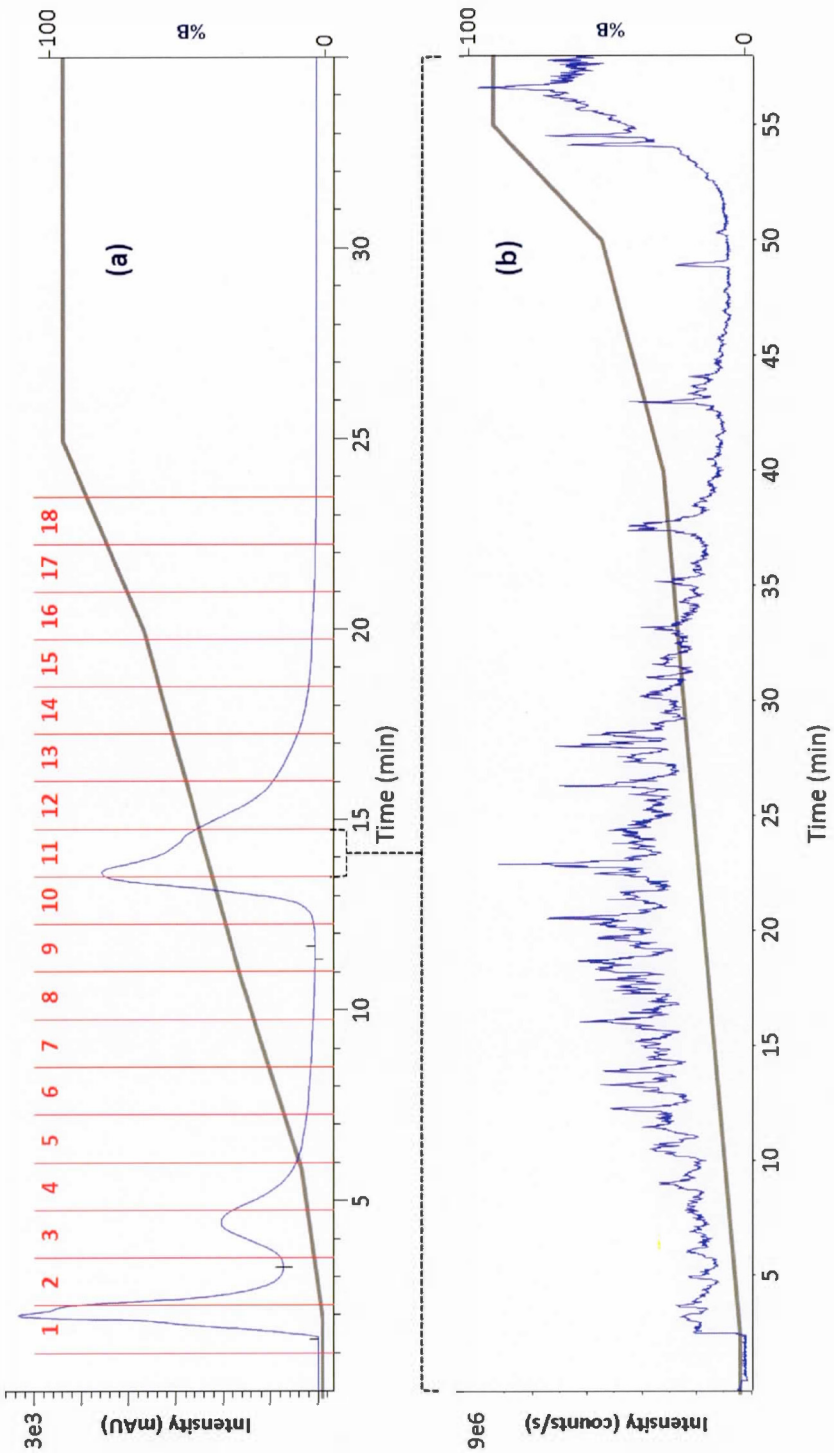


Figure 5.2 (a) UV trace at 220 nm from SCX chromatography of the APAP-treated rat liver microsomes digest. Vertical lines demonstrate fractionation intervals. (b) Representative total ion chromatogram from UHPLC-RP-LC-MS/MS analysis of SCX fraction #11 of the APAP-treated sample. The elution gradient is also shown in each trace.

A stand-alone java application (MassBox) was also developed to remove redundancies in peak lists and to filter peak pairs based on a delta-mass equal to the mass of APAP (considering +1 to +4 charge states). Jvenn interactive Venn diagram viewer (Bardou *et al.*, 2014) was used for overlap analysis and MultiQuant software was used for the final verification of modified peptides based on their absence in the control samples. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaíno *et al.*, 2014) via the PRIDE partner repository with the dataset identifier **PXD002590**.

5.4 Results and Discussion

A high resolution 2D-LC-MS/MS method was applied to APAP-treated and control (no drug) RLM incubations followed by extensive data mining to assess the presence of APAP-modified peptides. Data processing included (1) high-resolution peptide spectral matching, (2) differential analysis (following generic peak finding) between control and drug-treated sample, (3) detection of APAP-cysteine adduct-related diagnostic fragment ions, and (4) accurate mass difference filtering between treated and control samples.

5.4.1 Proteome Analysis of the Microsomal Samples

LC-MS/MS data from control and treated samples were subjected to generic proteomic searching to ascertain that our overall proteome coverage was consistent between all samples. An average number of 869±13 and 862±38 proteins (from 9661±145 and 9950±875 peptides) were identified at 1% FDR from the APAP-treated and control samples, respectively (n = 3). This comparison shows that results

for all samples were highly reproducible and therefore no bias would have been introduced simply by the treatment of liver microsomes with APAP.

5.4.2 Peptide Spectral Matching

An initial list of potential adducts was generated using ProteinPilot software where raw (*.wiff*) files from the SCX fractions were combined and searched against UniProtKB/Swiss-Prot (reviewed) database using peptide spectral matching (PSM). APAP modification was added as a custom modification (addition of $C_8H_8NO_2$ replacing H) on cysteine residues with a probability of 0.85. This increase in the probability of APAP-modification can lead to a relatively high bias towards the detection of APAP-modified peptide spectral matches and thus an increased number of potential false positives, needing to be subsequently removed using parallel complementary data mining strategies.

Using the custom PSM algorithm, 7666 distinct peptides (from 1% FDR proteins) were detected in the drug-treated sample, out of which 455 were identified to be potentially APAP-modified. Applying the same search parameters to the control sample, 134 of this list were rejected and were thus deemed false positives. The remaining 321 peptides were later subjected to an overlap analysis with the lists of putative APAP-modified peptides from other data mining workflows.

5.4.3 Statistical Differential Analysis

For differential analysis, MarkerView software was employed to generate peak lists from APAP-treated and control samples. After removing peak redundancies

within the range of 0.02 *m/z* and 0.5 min RT, 7491 peaks were considered potential hits and subjected to the overlap analysis with the lists of candidates from other data mining strategies.

5.4.4 Peak-Pair Finding

Peak lists (from MarkerView) from the treated and control samples were compared using MassBox Peak Picker to find peak-pairs with a delta-mass corresponding to the difference between APAP and IAM modification of a cysteine residue (149.0477 – 57.0215 = 92.0262 Da) for z = 1–4 within a mass accuracy of 25 ppm. Since APAP-modification should, in theory, increase hydrophobicity, peak pairs were considered valid only if the peak with the larger *m/z* value (in the treated sample) eluted later on the RP-LC column. The search yielded a total number of 181 peak pairs, which were later subjected to overlap analysis with results from other data mining strategies.

MassBox is an in-house custom-built open-source Java package for routine MS-based omics applications including molecular weight calculation, exact mass to *m/z* conversion, formula finding, redundancy removal, and peak-pair finding. The package has been deposited to SourceForge public source code repository and can be downloaded at http://massbox.sourceforge.net.

5.4.5 Differential Analysis Combined with Diagnostic Ion Screening

MetabolitePilot software was used to process chromatograms from each fraction to find peaks unique to APAP-treated samples for all LC-MS peaks eluting

from 3–60 min within *m/z* 400–1250. For a feature to be reported, the software needed to detect at least one of the four diagnostic (Sleno *et al.*, 2007) APAP-cysteine adduct-related product ions with an *m/z* tolerance of 25 ppm. As illustrated in Figure 5.3, these ions were detected from microsomal incubations with APAP and standard peptides. After the redundant features within the range of 0.02 *m/z* and 0.5 min RT were removed (using MassBox Redundancy Remover), a final list of 368 features of interest were subjected to overlap analysis with other data mining strategies.

5.4.6 Data Verification and Comparison of Data Mining Workflows

The final lists of potential APAP-modified peptides from each data treatment strategy were compared by *m/z*, RT and charge state (within 0.1 *m/z* and 1.5 min RT). As illustrated in Figure 5.4, out of the 7457 total features, 324 (4.3%) were detected by more than one data mining approach. Differential analysis with diagnostic ion screening (MetabolitePilot) resulted in the highest overlap (70%) with other strategies, most significantly with statistical differential analysis (MarkerView), which accounts for the highest number of candidates, but with only 5% of them shared with other workflows. Table 5.1 summarizes the number of putative APAP-modified peptides from each data analysis and the number of hits shared with other workflows.

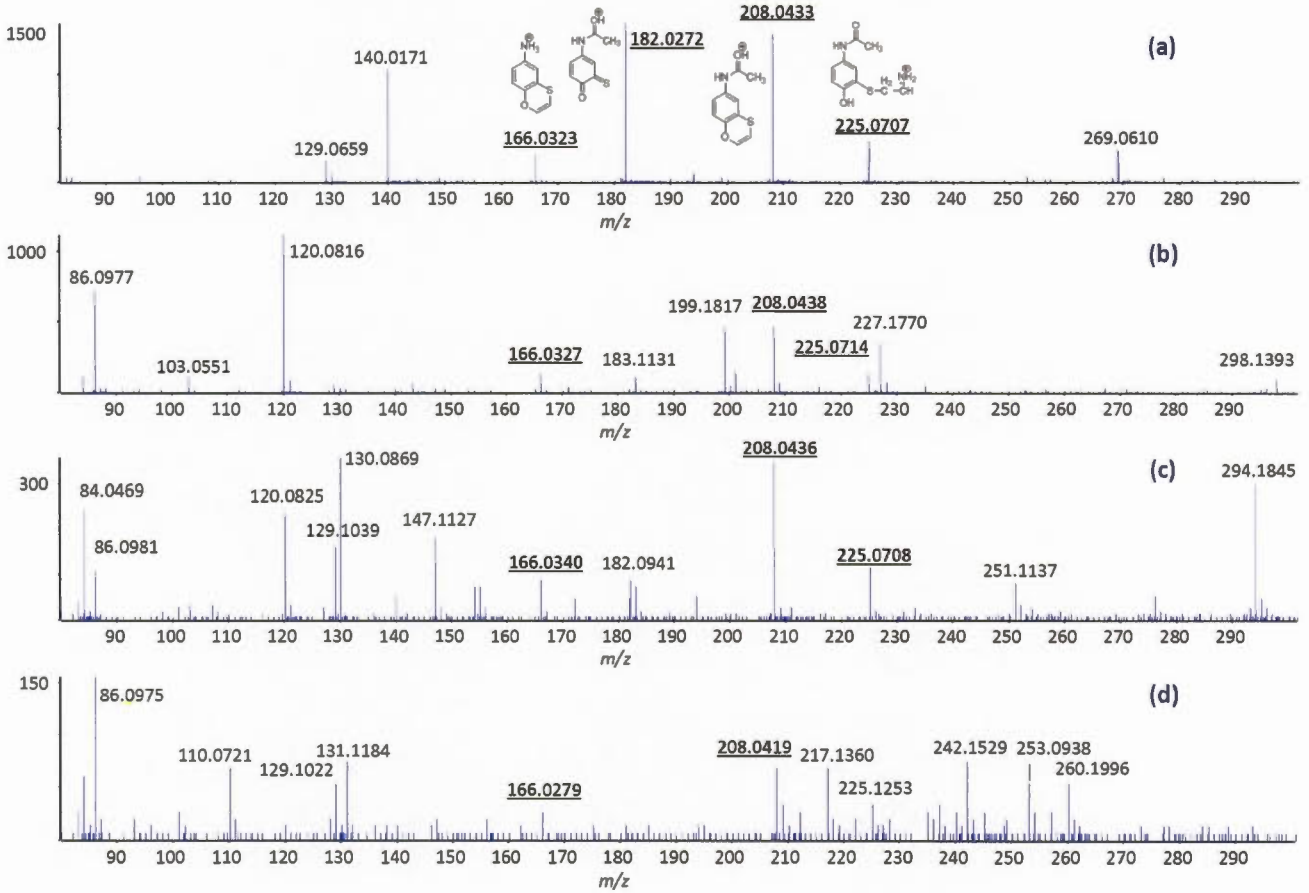


Figure 5.3 High-resolution MS/MS spectra (*m/z* 80–300 range) from APAP-modified (a) glutathione, (b) ILISDFGLCK, (c) QACLFK, and (d) LQQCPFEDHVKL peptides. The *m/z* values related to the APAP-cysteine adduct-related product ions are underlined

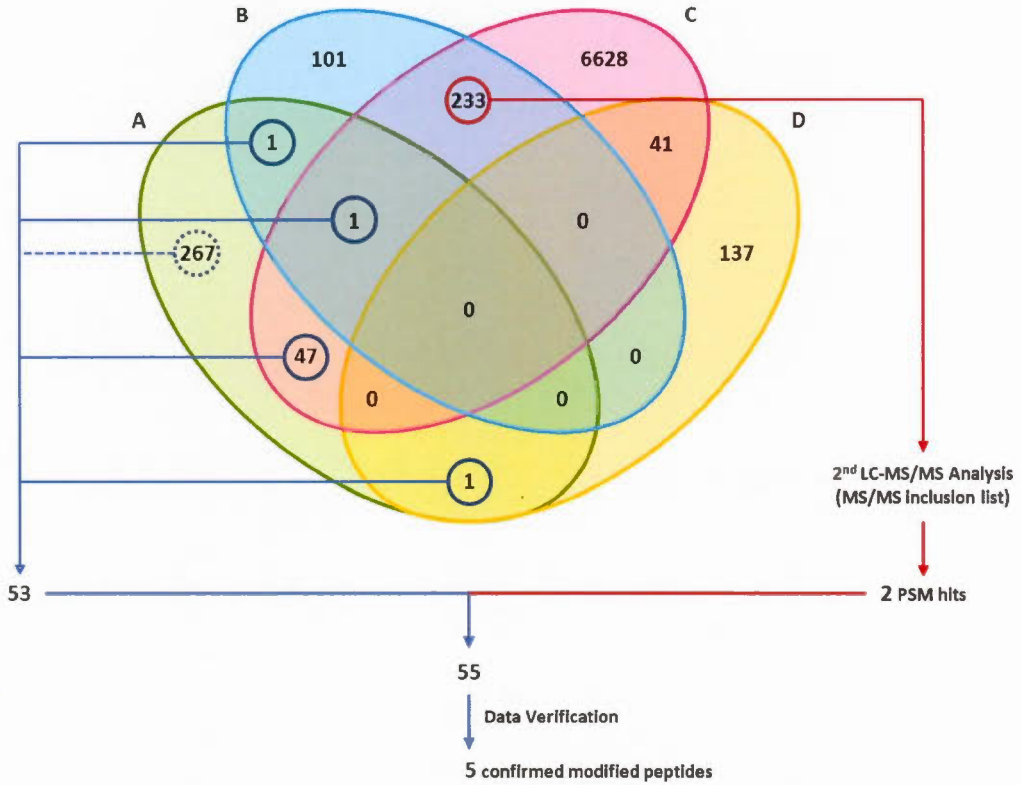


Figure 5.4 Data analysis workflow for the identification of APAP-modified peptides incorporating peptide spectral matching (A), differential analysis combined with diagnostic ion screening (separate SCX fractions) (B), statistical differential analysis based on LC-MS signal intensities (SCX fractions combined) (C), and peak-pair finding between control and treated samples (D). The number of putative APAP-modified peptides from each method (after redundancies were removed within 0.1 *m/z* and 1.5 min RT) and those confirmed after verification are also shown.

Table 5.1 Number of putative APAP-modified peptides detected by each data mining workflow and percent shared between different workflows

	Data mining workflow	Criteria	Software	# Candidates	# Shared	% Shared
A	**Peptide spectral matching**	MS/MS spectra Spectral confidence	ProteinPilot	317	50	16
B	**Differential analysis with diagnostic ion screening**	LC-MS signal ratio MS/MS spectra	MetabolitePilot	336	235	70
C	**Statistical differential analysis**	LC-MS signal ratio	MarkerView	6950	322	5
D	**Peak-pair finding**	Mass difference	MassBox	179	42	23

Features shared between the two differential analyses (233 peaks) were subjected to a second LC-MS/MS analysis using an inclusion list for a more targeted MS/MS acquisition to help characterize corresponding peptide sequences through PSM. Two additional putative APAP-modified peptides were found, which were then added to the list of previously characterized candidates from the PSM workflow that either were shared with other strategies or had a spectral confidence > 95%. The 41 hits which overlapped between peak pair-finding (MassBox) and differential analysis (MarkerView) data analysis workflows were eliminated based on being present in the control samples upon further verification (not included in inclusion list). The final list of 55 peptides was then visually inspected in terms of accurate mass (within 25 ppm), charge state and presence in the control samples. More specifically, potential hits were reduced to 46 upon mass accuracy verification (within 25 ppm), and then to 33, after filtering based on isotope pattern and charge state.

Five peptides were ultimately confirmed as APAP-modified based on their high-resolution LC-MS/MS data and being unique to the drug-treated sample. Table 5.2 lists the identified adducted peptides, their representative proteins and sites of APAP modification. Extracted ion chromatograms, high-resolution MS and MS/MS spectra from the confirmed APAP-modified peptides related to these proteins can be found in Appendix A.

During data verification, several highly confident (peptide confidence 99%) false positives were removed. For example, LKVFVDLLPAAQ**C**TQFINQLLGVVPLST, Asn-deamidated version of VFANPED**C**AGFGKGENAK and *N*-terminal carbamoylated LAGIVSWGDA**C**GAP were all found to be APAP-modified within 25 ppm by both PSM and statistical differential analysis. However, the first peptide had a wrong charge state; the second was not a mono-isotopic peak, and the third was found to also be present in the control, and thus all three were rejected.

Table 5.2 List of the identified APAP adducts of RLM proteins

APAP-Modified Peptide	Target Protein	Accession	Modification Site	Detection Strategy
FIDLIPTNLPHAVTCDIK	Cytochrome P450 2C6	CP2C6_RAT	Cys-372	PSM (conf. > 95%)
GKEGAHAPCASE	MAP kinase-activating death domain protein	MADD_RAT	Cys-125	PSM DA (LC-MS)
VFANPEDCAGFGK	Microsomal glutathione S-transferase 1	MGST1_RAT	Cys-50 (active site)	PSM DA (MS/MS)
VFANPEDCAGFGKGENAK				PSM DA (LC-MS) DA (MS/MS)
MPFTSSCLII	NADH-ubiquinone oxidoreductase chain 5	NU5M_RAT	Cys-371 (TMD)	PSM DA (MS/MS)

DA (LC-MS): statistical differential analysis based on LC-MS signal intensities; DA (MS/MS): differential analysis combined with diagnostic ion screening; PSM: peptide spectral matching; TMD: transmembrane domain

Multiple comparison procedures generally help reduce false discovery rate thus increasing the validity of results (Bate and Clark, 2014). The probability of the same feature being detected by more than one data processing strategy decreases the risk of that feature being a false positive. In the present study, initial hits were considered a "potential" candidate only if high-quality MS/MS data supported the peptide spectral match (PSM confidence > 95%) or they were detected by at least one additional detection method other than the PSM workflow, which is often the basis of MS-based target protein identification studies (Qiu *et al.*, 1998; Jan *et al.*, 2014; Yang *et al.*, 2014). Although no real hits resulted from delta mass filtering, this remains an interesting tool to search for specific modifications in complex proteomic samples. Also, the inclusion list of 233 ions, only 2 PSM hits were added to the list of previously characterized peptides resulting from the initial IDA method employed. This fact reassured us that the IDA parameters were optimized as to trigger a very large majority of interesting ions in these samples even considering their low abundance relative to the other peptides present in these complex samples. This model had been optimized previously for proteome coverage in rat liver microsomal samples (Golizeh *et al.*, 2015c).

To further confirm the structure of the detected APAP-modified peptides, synthetic standards of target peptides were purchased, subjected to microsomal incubations with APAP, and analyzed using the same LC-MS/MS method. Co-elution of the extracted ion chromatograms, as well as MS and MS/MS spectral matching between APAP-modified synthetic peptides and those found from digested rat liver microsomes proved that these modification sites were accurately characterized (see Appendix B).

5.4.7 APAP Covalent Binding Target Proteins

Over 30 proteins have been found to be covalently modified by APAP *in vivo* in mouse, rat, and human (Qiu *et al.*, 1998; Jan *et al.*, 2014; Streeter *et al.*, 1984; Axworthy *et al.*, 1988; Dietze *et al.*, 1997; Cohen *et al.*, 1997; James *et al.*, 2003; Copple *et al.*, 2008; Damsten *et al.*, 2007; Senter *et al.*, 2002; Wendel and Cikryt, 1981; Zhou *et al.*, 1995). A comprehensive list of these proteins can be obtained from the reactive metabolite target protein database (TPDB) (Hanzlik *et al.*, 2007). In the present study, four proteins were identified as APAP-modified, three of which were not previously reported as *in vivo* targets of drug-induced covalent modification.

Microsomal glutathione *S*-transferase 1 (MGST1) is a multi-pass integral membrane protein in ER and mitochondria outer membrane, which is involved in conjugation of GSH to a wide number of exogenous and endogenous hydrophobic electrophiles and is a crucial part of the cellular response to drugs, toxins, and other xenobiotics. It is known that the Cys-50 residue of this protein acts as a switch for enzyme activation if being attacked by electrophilic species including NAPQI (Shin *et al.*, 2007). In this study, two Cys-50-containing tryptic peptides were APAP-modified at 99% spectral confidence. VFANPED**C**AGFGK (*m/z* 752.3323, 21.9 min, 2+) and VFANPED**C**AGFGKGENAK (*m/z* 668.3036, 19.3 min, 3+) were detected in all three technical replicates by both peptide spectral matching and differential analysis. The latter also showed an APAP-cysteine adduct-related diagnostic ion at *m/z* 208.04. Excessive modification of MGST1 by NAPQI can accelerate depletion of the protective GSH reservoir rendering other potential cellular targets vulnerable to covalent additions. The primary structure of MGST1 is highly homologous in rat, mouse and human and it is possible that a similar modification occurs in APAP-induced human liver injury.

Cytochrome P450 2C6 (CP2C6) is a liver microsomal heme-thiolate monooxygenase involved in NADPH-dependent electron transport and oxidizes a variety of hydrophobic compounds, including steroids, fatty acids, and xenobiotics. The CP2C subfamily is the most complex subfamily of the P450s found in human and animal species with several different isoforms, and is involved in the metabolism of a large number of drugs on the market. The human ortholog of rat CP2C6 is CP2C9, which is the major form of the 2C subfamily and accounts for 60% of total human CP2C. The mouse counterpart of this protein is CP2C29, which is also the most abundant CP2C in this species (Martignoni *et al.*, 2006). These results demonstrate that the highly conserved Cys-372 of this protein has been modified by APAP. The tryptic peptide containing this modification site, FIDLIPTNLPHAVTCDIK (*m/z* 720.3784, 33.7 min, 3+), was detected at 99% spectral confidence in all three technical replicates, and was only found by peptide spectral matching. Although strongly involved in drug metabolism, CP2C6 has not been reported as a covalent target before. The only members of the CP2C subfamily reported to be modified by reactive metabolites are 2C7 and 2C11 that were both identified in rat *in vivo* in relation to teucrin A (Druckova *et al.*, 2007) and diclofenac (Shen *et al.*, 1997), respectively. Moreover, in addition to liver, CP2C9 and its orthologs have also been detected in several other tissues such as kidney, testes, adrenal gland, prostate, ovary and small intestine (Martignoni *et al.*, 2006).

NADH-ubiquinone oxidoreductase chain 5 (NU5M) is a core subunit of the respiratory chain NADH dehydrogenase (Complex I) that translocates electrons and protons across the inner mitochondrial membrane helping to maintain the electrochemical potential difference required to produce ATP. This enzyme is the first and largest component of the electron transport chain and, due to its structural, functional and regulatory complexity, is particularly vulnerable to genetic and sporadic pathological factors (Papa *et al.*, 2012). Complex I couples the oxidation of NADH to the reduction of ubiquinone. During this process, electrons can escape from

the complex and react with ambient O_2 to produce O_2^- and the subsequent reactive oxygen species (ROS) (Koopman *et al.*, 2010). Covalent modification of this protein can impact the efficiency of the enzymatic process and potentially play a role in the initial mitochondrial oxidative stress triggered by protein adduct formation observed in APAP hepatotoxicity (Jaeschke *et al.*, 2003; Jiang *et al.*, 2015). From the present study, Cys-371 of NU5M was identified as APAP-modified. The non-tryptic peptide MPFTSSCLII (*m/z* 630.8038, 5.3 min, 2+) was detected by both peptide spectral matching and differential analysis with an APAP-cysteine adduct-related diagnostic ion at *m/z* 182.03. Cys-371 lies on one of the 15 transmembrane domains of the Complex I subunit and, although conserved in mouse, this cysteine is replaced with a serine residue in the human ortholog, and thus cannot explain the APAP-induced mitochondrial dysfunction seen in human hepatocytes (Jiang *et al.*, 2015; Jaeschke *et al.*, 2012). Although microsomal fractions should in theory be composed mainly of proteins from the ER membrane, they are often not free of cell debris, nuclei and mitochondria. For instance, in our previous analyses of rat, mouse and human liver microsomes (Golizeh *et al.*, 2015c), 13.3, 13.6 and 12.4% of the proteins were identified as mitochondrial, respectively. Therefore, the identification of NU5M, a mitochondrial membrane protein, as a covalent target of APAP in the studied samples should not be surprising.

Mitogen-activated protein kinase (MAPK)–activating death domain protein (MADD) plays a crucial role in regulating cell proliferation, survival and death through alternative mRNA splicing. MADD converts GDP-bound inactive Rab3 (A, C and D) to its GTP-bound active form (Del Villar and Miller, 2004). This protein was found to be APAP-modified on Cys-125. The non-tryptic peptide GKEGAHAPCASE (*m/z* 653.2800, 8.4 min, 2+) containing this modification site was detected by the peptide spectral matching and statistical data analysis strategies. Cys-125 is located on the region that connects uDENN and DENN domains. The DENN (differentially expressed in normal versus neoplastic) domain is thought to be

involved in GTP/GDP exchange activity and thus essential to the function of the protein (Levivier *et al.*, 2001). This cysteine is replaced by a glycine residue in mouse, but is conserved in the human ortholog. Jaeschke *et al.* (2013) proposed that APAP-induced mitochondrial dysfunction and oxidative stress activates apoptosis signal-regulating kinase 1 (ASK1) and c-Jun *N*-terminal kinase (JNK) leading to enhanced ROS formation. The resulting damage to the mitochondrial membrane leads to the release of mitochondrial endonucleases (AIF, EndoG) causing nuclear DNA fragmentation and eventually cell necrosis. MADD is known to also promote JNK activation through a series of death domain interactions enabling TRAF2–mediated activation of ASK1 (Del Villar and Miller, 2004) therefore, its APAP-modification at its DENN domain region can potentially play a role in the suggested cell death mechanism. Figure 5.5 places each of the identified APAP-modified proteins in the grand scheme of previously suggested mechanisms for APAP-induced hepatotoxicity (Jaeschke *et al.*, 2013).

Generally, rats are deemed to be more resistant to APAP toxicity than mice, and although the overall APAP metabolism is similar in both species, the extent of drug-induced oxidative stress and hepatocyte necrosis due to mitochondrial dysfunction are significantly lower in rats (McGill *et al.*, 2012). For this reason, very few reports exist on the study of APAP-induced hepatotoxicity in rats. However, acetaminophen toxicity in rat has been the subject of several recent studies (Mahmoud *et al.*, 2015; Mast *et al.*, 2015; Yao *et al.*, 2015; İçer *et al.*, 2016). The APAP-modification of rat proteins identified in this study will be compared to other species in the future.

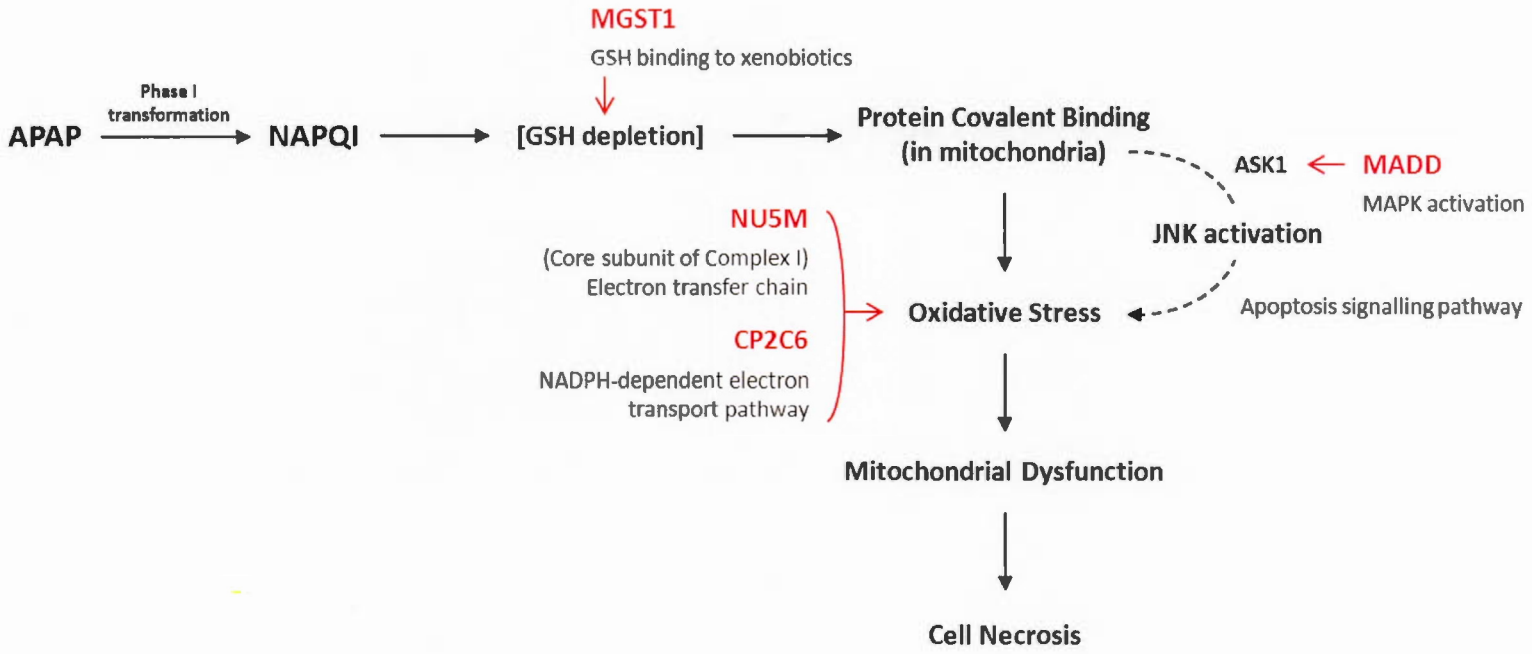


Figure 5.5 Putative role of the identified modified proteins in the previously suggested mechanism for APAP-induced hepatotoxicity. MGST1, MADD, NU5M, and CP2C6 are all involved in the regulation of the oxidative stress and their covalent modification could therefore potentially affect their biological functions, altering critical cellular pathways that can cause or augment cytotoxic consequences.

5.5 Conclusions

An approach was developed to identify protein targets of APAP covalent binding from rat liver microsomes using 2D-LC-HR-MS/MS and multiple complementary data mining workflows. SCX fractionation was employed to decrease sample complexity and increase the potential for detecting modified peptides. The detailed data processing strategy was aimed at removing any false positives, an important drawback of generic methods employed for identification of target proteins using LC-MS/MS. Four proteins were confidently identified to be adducted by the reactive metabolite of APAP (NAPQI), namely MGST1, CP2C6, NU5M and MADD. These proteins are all known to be involved in critical biological pathways that govern cell survival under xenobiotic-induced oxidative stress and their modification can be potentially linked to the mechanisms proposed for APAP-induced hepatotoxicity.

5.6 Acknowledgements

M. G. thanks GRASP (*Groupe de Recherche Axé sur la Structure des Protéines*, FRSQ) for a graduate student scholarship.

5.7 Supporting Information

High-resolution UHPLC-RP-MS/MS data from the APAP-modified peptides identified in rat liver microsomal incubations (Appendix A); comparison of the MS and MS/MS spectra from the APAP-modified peptides identified in rat liver microsomes to those from the trapping experiments using synthetic peptides (Appendix B).

CHAPTER SIX

COVALENT BINDING OF 4-HYDROXYNONENAL TO MATRIX METALLOPROTEINASE 13 STUDIED BY LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY

Makan Golizeh, Jamilah Abusarah, Mohamed Benderdour and Lekha Sleno

Published in **Chemical Research in Toxicology 2014, 27, 1556-1565**. Supporting information included in Appendix C (also available online at doi:10.1021/tx5002095).

Covalent modification of proteins by reactive metabolites is not limited to xenobiotics. Numerous endogenous metabolites are chemically reactive and readily undergo covalent reactions with macromolecules, such as proteins, to form stable adducts affecting the biological function of the substrate, potentially inducing toxicity. In this chapter, an analytical workflow was developed for the detection of covalent adduction of matrix metalloproteinase 13 by 4-hydroxynonenal, an endogenous reactive metabolite, believed to be involved in the pathogenesis of osteoarthritis. The method was first applied to *in vitro* incubations and then optimized for the analysis of clinically-relevant samples.

Makan Golizeh, Jamilah Abusarah, Mohamed Benderdour and Lekha Sleno are co-authors of this article. Makan Golizeh conducted the literature survey, prepared experimental protocols, conducted the experiments, processed the data, and prepared the original manuscript. Jamilah Abusarah participated in collection and preparation of the biological samples. Mohamed Benderdour proposed the original idea of the project, provided biological samples, and contributed to the interpretation of results. Professor Lekha Sleno supported and supervised the project, verified data

analysis and interpretation of the results, revised and finalized the manuscript. The second and third authors of this article were, at the time of publication, affiliated with Orthopaedic Research Laboratory of *Hôpital du Sacré-Cœur*, Université de Montréal, Montreal, Quebec.

6.1 Abstract

Osteoarthritis (OA) is caused by the degradation of articular cartilage and affects approximately 80% of people over the age of 65. Matrix metalloproteinases (MMPs) belong to a group of zinc endopeptidases that degrade extracellular matrix (ECM) proteins in cartilage. MMP-13, also known as collagenase 3, cleaves type II collagen more rapidly than other MMPs and therefore is an important target for the treatment of OA. The lipid peroxidation product 4-hydroxy-2-(*E*)-nonenal (HNE), generated under oxidative stress, is known to play a crucial role in cartilage degradation, however the mechanism is not yet fully understood. An approach has been developed to monitor HNE modification sites by incubating rhMMP-13 ± HNE *in vitro* followed by analysis of tryptic digests by UHPLC coupled to high resolution (HR) quadrupole-time of flight (QqTOF) tandem mass spectrometry (MS/MS). The analysis elucidated several covalently-modified histidine and cysteine residues. The reaction was monitored using different HNE concentrations and incubation times. A targeted assay, using multiple-reaction monitoring (MRM), was then optimized to increase the sensitivity of detecting these modification sites in biological samples. HNE-related covalent modifications of MMP-13 were confirmed in enriched extracts from interleukin 1β-activated chondrocytes from OA patients using high resolution MS/MS and MRM analysis.

6.2 Introduction

Biologically-active electrophiles are a diverse set of compounds that are produced in living organisms and tend to covalently bind to nucleophilic centers present in biomolecules such as nucleic acids and proteins, and therefore can interfere with a wide range of biological functions (Gersch *et al.*, 2012). α,β-Unsaturated carbonyls represent an important group of such reactive species and have been shown to exhibit important toxicity (Kalgutkar, 2011). 4-Hydroxy-2-(*E*)-nonenal (HNE) is an α,β-unsaturated aldehyde and reactive oxygen species (ROS) endogenously generated by peroxidation and breakdown of both ω-3 and ω-6 polyunsaturated fatty acids (PUFA) under oxidative stress. This electrophilic species has been linked to a number of adverse biological effects, affecting cell proliferation and gene expression through carbonylation of essential biomolecules (Esterbauer *et al.*, 1991). HNE is highly reactive and has the ability to covalently bind to carbohydrates, lipids, nucleic acids, amino acids and proteins as well as low molecular weight metabolites, such as glutathione. Covalent binding of HNE to protein yields adducts which can alter the protein's structure and function causing cell damage as well as adverse immune responses (Niemelä, 1999). In a recent review article on protein damage (Chondrogianni *et al.*, 2014), the mechanism of this impairment is classified under "secondary oxidative protein modifications" occurring when proteins are indirectly modified through secondary reactions with by-products of oxidative stress. This is in contrast to other types of ROS, such as free radicals, that directly bind to proteins. The addition of HNE onto proteins often occurs on free thiol groups of cysteine residues as well as the imidazole ring nitrogen of histidine, forming stable adducts through a 1,4-addition via the Michael reaction. However, the ε-amino group of lysine, guanidine nitrogen of arginine and the *N*-terminal amine may also generate less stable and potentially reversible HNE adducts via Schiff base formation (Grimsrud *et al.*, 2008). Although some studies show that the reactivity of these amino acids to HNE occurs in the following order: Cys > His > Lys > Arg (Sayre *et*

al., 2006; Doorn and Petersen, 2003), preference of the reaction sites also depends on the protein's structure (Uchida and Stadtman, 1992). It has been shown that histidine residues are often the primary targets for HNE (Roe *et al.*, 2007). Michael adducts represent 99% of HNE protein modifications, while Schiff base formation is much less prevalent, even in the presence of excess HNE (Bruenner *et al.*, 1995).

Matrix metalloproteinases (MMPs) are a family of zinc endopeptidases in humans that enable extracellular matrix (ECM) breakdown under normal and disease conditions (Kevorkian *et al.*, 2004). Unusual expression or activity of MMPs is known to be associated with several disorders such as metastasis, angiogenesis, and cardiovascular diseases, as well as rheumatoid arthritis (RA) and osteoarthritis (OA) (Engel *et al.*, 2005). It has been reported that HNE is involved in accelerated production and increased activity of several MMPs including MMP-1 (Akiba *et al.*, 2006; Chapple *et al.*, 2013; Zamara *et al.*, 2004), MMP-2 (Chapple *et al.*, 2013; Zamara *et al.*, 2004; Lee *et al.*, 2008), MMP-9 (Lee *et al.*, 2010; Minami *et al.*, 2009), and MMP-13 (Morquette *et al.*, 2006). It is suggested that this reactive species degrades cartilage through either oxidation of ECM macromolecules or covalent binding to MMPs resulting in their activation via a "cysteine-switch" mechanism (Gu *et al.*, 2002; Fu *et al.*, 2001). MMP-13, or collagenase 3, is produced by the chondrocyte and preferentially cleaves type II collagen and is essential in the development of degenerative joint disease (Engel *et al.*, 2005). Collagenases, such as MMP-13, are the only mammalian proteinases that can specifically degrade fibrillar collagens at neutral pH, and thus play a crucial role in cartilage homeostasis (Kevorkian *et al.*, 2004). The nature of the interaction between HNE and MMP-13 and its biological relevance with collagen degeneration, if better understood, could present a new avenue for the treatment of OA.

Mass spectrometry (MS)-based proteomic analysis is one of the most commonly used analytical techniques for studying protein adducts and covalent

binding (Liebler, 2008). The analytical workflow is usually based on a bottom-up approach, where proteins are digested with proteases, and peptides are then separated (by electrophoresis or chromatography) prior to peptide sequencing by tandem mass spectrometry, most often using high-resolution mass analyzers (Khan *et al.*, 2011). On the other hand, triple-quadrupole (QqQ) systems have lower resolving power but offer higher sensitivity in targeted assays and thus are often applied to complex biological samples, where specific peptides are monitored using multiple-reaction monitoring (MRM) with the capacity for the simultaneous analysis of hundreds of peptides (Walther and Mann, 2010). Whereas high-resolution mass spectrometry (HR-MS) enables comprehensive profiling and structural analysis, MRM assays are reliable tools for targeted analyses (Liebler and Zimmerman, 2013).

In the present study, a proteomics approach was employed using UHPLC coupled to HR-MS/MS to characterize HNE-related covalent modifications of MMP-13, and to screen the extent of protein covalent binding under different treatment conditions, such as time of reaction and HNE concentration. This method was subsequently applied to confirm these modifications in an enriched chondrocyte extract from patients with OA, in conjunction with a targeted LC-MRM analysis to detect the corresponding HNE-modified peptides with higher sensitivity.

6.3 Experimental

6.3.1 Materials

Recombinant human matrix metalloproteinase 13 was purchased from R&D Systems (Minneapolis, MN), sequencing-grade trypsin was from Promega (Madison, WI) and 4-hydroxy-2-(*E*)-nonenal was obtained from Cayman Chemicals (Ann Arbor, MI). Dulbecco's modified Eagle's medium, fetal bovine serum, penicillin and

streptomycin were from Life Technologies (Burlington, ON, Canada) whereas interleukin-1β was from R&D System (Minneapolis, MN). *Streptomyces griseus* pronase, type IV collagenase and all other chemical reagents (analytical grade) were purchased from Sigma-Aldrich (St. Louis, MO). HPLC-grade solvents were acquired from EMD Chemicals (Gibbstown, NJ). All aqueous solutions were prepared using nanopure water purified by a Millipore Synergy UV system (Billerica, MA).

6.3.2 Incubation of Recombinant Human MMP-13 with HNE

Recombinant human MMP-13 (2.5 μg) was incubated in the presence of HNE at 0, 8, 40, 200, 1000 and 2400 μM (total volume of 150 μl in 100 mM phosphate buffer pH 7.4) for 2 h at 37°C, followed by an additional 30 min with cysteine (final concentration 0.4 mM) to deactivate the excess HNE. For the time-course experiment, MMP-13 was incubated with HNE at 200 μM for up to two hours.

6.3.3 Specimen Selection, Chondrocyte Culture and Treatment

Post-surgery discarded human OA articular cartilage was obtained from OA patients (aged 67 ± 9 years, $n = 4$) and informed consent was obtained for the use of tissues. All patients were evaluated by rheumatologists who followed American College of Rheumatology criteria (Altman *et al.*, 1986). The protocols and use of human tissues were approved by the Research Ethics Board of *Hôpital du Sacré-Cœur de Montréal*. OA knee cartilage specimens were spliced and rinsed, and chondrocytes were extracted by sequential enzymatic digestion, as described previously (Morquette *et al.*, 2006). Cartilage samples were digested with pronase (1 mg/ml, 1 h, 37°C) and type IV collagenase (2 mg/ml, 6 h, 37°C) in Dulbecco's

modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 100 units/ml of penicillin and 100 μg/ml of streptomycin. The cells were seeded at high density in culture flasks at 37°C in a humidified atmosphere of 5% CO_2/95% air until they were confluent. Chondrocytes were then incubated (48 h) in DMEM (10 ml) with 0.1% FBS in the presence of 1 ng/ml interleukin-1beta (IL-1β). The medium was then collected and subjected to either ion exchange chromatography or immunoprecipitation (IP).

6.3.4 Immunoprecipitation

The culture medium was passed over a column containing aldehyde-activated beaded agarose (Pierce, Rockford, IL) to which mouse anti-human MMP-13 antibodies were covalently attached through primary amines. After washing with PBS, human MMP-13 was eluted with 100 mM glycine buffer (pH 3). The eluate was dialyzed overnight at 4°C against PBS and concentrated using Amicon 30 kDa ultra-centrifugal filter devices (EMD Millipore, Etobicoke, ON, Canada). Purified MMP-13 was quantified with a commercial ELISA kit from R&D Systems and HNE/MMP-13 adduct was detected by Western blot using anti-HNE antibody as primary antibody. MMP-13 concentration was 2.5 and 1 μg/ml before and after IP, respectively.

6.3.5 Protein Fractionation by Ion Exchange Chromatography

The culture media was concentrated using Amicon 10 kDa ultra-centrifugal filter devices to remove low-MW impurities. The sample (~20 μl) was then diluted in 100 μl of buffer A and injected (100 μl, 1.25 μg protein) onto a PolyCATWAX

200 × 2.1 mm column with 5 µm (1000 Å) particles (PolyLC, Columbia, MD) using an Agilent 1200 series HPLC (Agilent Technologies, Palo Alto, CA) equipped with a binary pump, degasser, diode array detector and fraction collector. Fractionation was performed at 200 µl/min using a gradient of 100% A held for 1 min, up to 8% B at 8 min, 85% B at 9.5 min, then to 100% B at 12.5 min and held for an additional 17.5 min, where buffers A and B were 10 mM and 800 mM ammonium acetate in 20% acetonitrile (pH 7.2), respectively. UV absorbance was monitored at 280 nm. Eight fractions were collected at the following intervals: 2-4, 4-6, 6-10, 10-14, 14-18, 18-22, 22-26 and 26-30 min. Collected fractions were evaporated to dryness under vacuum (Thermo Fisher Scientific Universal Vacuum System, Asheville, NC) and then reconstituted in 100 mM ammonium bicarbonate (ABC) pH 8.5 (100 µl) prior to digestion.

6.3.6 Protein Digestion

Reductive alkylation was performed with dithiothreitol (DTT) (10 min, 25°C) and iodoacetamide (IAM) (30 min, 37°C, in the dark) at a final concentration of 10 and 40 mM, respectively. Samples (containing 0.1-1 µg protein) were diluted in ABC (200 µl) and digested using trypsin at a 1:15 (w/w) enzyme : protein ratio for 16 h at 37°C. Samples were cleaned-up using solid-phase extraction (SPE) on OASIS HLB (Waters, Milford, MA) cartridges (1 cc, 30 mg) and eluted with 100% methanol (1 ml). Dried extracts were reconstituted in 10% acetonitrile (50 µl) prior to analysis.

6.3.7 LC-HR-MS/MS Analysis

Samples were injected (20 µl) onto a Kinetex XB-C_{18} 2.1 × 100 mm column, with solid core 1.7 µm particles (100 Å) (Phenomenex, Torrance, CA) using a Nexera UHPLC system (Shimadzu, Columbia, MD). Reversed-phase liquid chromatography was performed using water (A) and acetonitrile (B), both containing 0.1% formic acid, with a gradient of 5-30% B in 15 min, to 40% B at 24 min, followed by a sharp increase to 80% B at 25 min and held for an additional 3 min prior to re-equilibration of the column, with a flow rate of 300 µl/min and column temperature of 40°C. MS and MS/MS spectra were collected on a hybrid quadrupole-time-of-flight (QqTOF) TripleTOF 5600 mass spectrometer (AB Sciex, Concord, ON, Canada) equipped with a DuoSpray ion source in positive ion mode set at 5 kV source voltage, 500°C source temperature and 50 psi for GS1/GS2 gas flows, with a declustering potential of 80 V. Data were acquired in IDA (information-dependent acquisition) mode, with a survey TOF-MS acquisition from *m/z* 140-1250, followed by MS/MS on the ten most intense ions (+1 to +4 charge states) from *m/z* 100-1500 with dynamic background subtraction (DBS) with a total cycle time of 1.3 s. MS/MS of precursor ions was performed at a collision energy of 30 ± 10 V.

6.3.8 LC-MRM Analysis

Samples were injected (5 µl) and eluted using the same chromatographic method as above. Multiple-reaction monitoring (MRM) was performed on a hybrid quadrupole-linear ion trap QTRAP 5500 mass spectrometer (AB Sciex, Concord, ON, Canada) equipped with a Turbo V ion source in positive mode. Nine peptides were monitored in modified and unmodified forms (different charge states) with 87 MRM transitions in total, with optimized collision energies and a total cycle time of

1.2 s. Ion spray voltage was set to 5 kV with a temperature of 450°C, declustering potential of 80 V and GS1/GS2 at 50 psi.

6.3.9 Data Processing

LC-HR-MS/MS data were processed using ProteinPilot (version 4.1) and MetabolitePilot (version 1.5) software for finding HNE-related modifications. MS/MS files were searched, with no enzyme specification, against the FASTA file obtained from the MMP13_HUMAN entry (accession number P45452) of UniProt knowledgebase (last modified April 3, 2013, version 138) by ProteinPilot software using Paragon search algorithm (Shilov *et al.*, 2007). For cell culture samples, the MS/MS data were searched against the (reviewed) human subset of uniprot_sprot.FASTA from UniProt. Proteins were identified with a 1% global false discovery rate (FDR) with a target-decoy database search algorithm (Tang *et al.*, 2008) using ProteinPilot Descriptive Statistics Template (version 3.001p) (http://www.absciex.com/PDST). Mass tolerance was 0.05 Da on precursor ions and 0.1 Da on fragments and the search was performed for +2 to +4 charge states. In order to find potential HNE adducts, the Paragon protein search algorithm was changed to include higher probability (0.99) for HNE adducts on cysteine, lysine and histidine. MetabolitePilot software was employed to compare HNE-treated sample with the control (-HNE) to find unique peaks in the treated sample, while PeakView software (version 1.2) was used to visually inspect and verify extracted ion chromatograms (XICs), MS and MS/MS spectra of the potential adducts. Given the accurate mass and retention times (RT) of the modified and unmodified peptides, MultiQuant software (version 2.1) was employed for peak area integration.

6.4 Results and Discussion

6.4.1 Covalent Modification of MMP-13 by HNE

In order to determine possible HNE modification sites on MMP-13, recombinant human matrix metalloproteinase 13 (MMP-13) was incubated with HNE *in vitro* and the digested protein was then subjected to LC-HR-MS/MS analysis. Based on high resolution MS and MS/MS data, MMP-13 was identified with over 99% sequence coverage in the control and over 97% in the HNE-treated sample. The protein sequence was identified based on a total number of 276 and 147 peptide features with over 95% confidence for the control and treated samples, respectively. Peptide features includes all the tryptic and non-tryptic peptide sequences with different post-translational modifications and charge states. Since each peptide sequence may be detected multiple times as different features, this number is higher than the number of unique protein sequences, known as the proteotypic peptides.

Two different but complementary strategies were used to examine HNE-associated covalent modifications on the MMP-13 sequence using ProteinPilot and MetabolitePilot software packages. The Paragon search algorithm was slightly altered to allow ProteinPilot to detect HNE modifications on histidine, cysteine and lysine residues with higher probabilities, leading to the identification of several HNE-modified histidine- and cysteine-containing peptides. Since there is a possibility that some HNE-modified peptides could be neglected by the search engine as a result of their low-confidence or missing MS/MS data, the HNE-treated sample was also compared to the control (-HNE) in MetabolitePilot software to find unique features and a complementary list was compiled after visual inspection of peaks. Potential adducts found via this search were then added to the list of modified peptides found with ProteinPilot.

Once a final list was compiled, the HNE-treated sample was verified for the unique peaks using PeakView software. Extracted ion chromatograms (XICs), as well as MS and MS/MS spectra, were reported for all unique potential adducts for further investigation. With the final confirmed list of interesting peaks, each potential adduct was used to calculate the theoretical non-adducted peptide. This final list of adducted and non-adducted peptides was then used to generate an inclusion list (as IDA criteria) for a second round of LC-MS/MS analysis through which each pair of adducted/non-adducted peaks were compared in terms of accurate mass, retention time shift and MS/MS behavior. Covalent modification by HNE was detected on seven histidines (His-48, His-187, His-251, His-312, His-340, His-343, and His-387) as well as two cysteines (Cys-96 and Cys-284). The sequence of human MMP-13 and the location of the detected HNE modifications are illustrated in Figure 6.1.

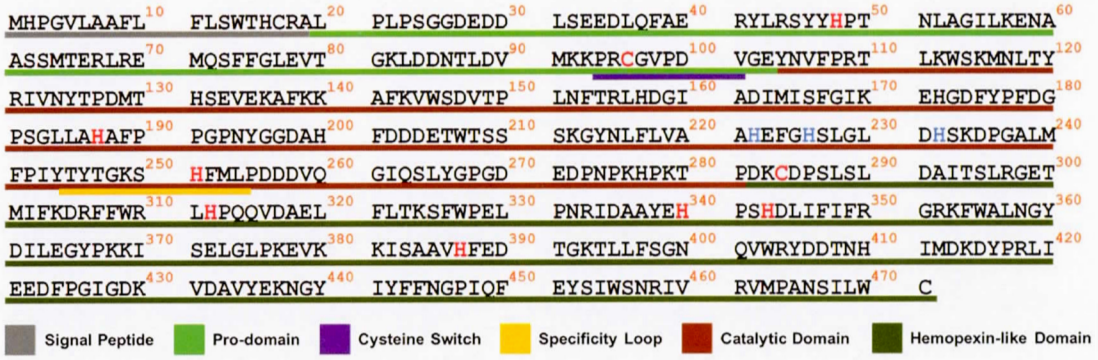


Figure 6.1 Sequence annotation of human MMP-13 (UniProt accession P45452) and the identified HNE modification sites (catalytic sites colored in blue and HNE-modified sites in red)

The chemistry of HNE modification involves a 1,4 Michael addition of the imidazole ring's non-acidic nitrogen to the gamma carbon (C-3) of HNE, followed by an intramolecular cyclization to form an α-tetrahydrofuryl alcohol (Uchida and

Stadtman, 1992). In this study, a subsequent dehydration of the alcohol into a substituted dihydrofuran was also detected in the case of histidine modification. It is presumed that the thiol group's sulfur of cysteine would also react with HNE in a similar way; however, additional dehydration was not detected for any of the modified cysteines (Figure 6.2). Considering the formerly reported (Bruenner *et al.*, 1995) 99:1 ratio of HNE Michael adduct formation over its Schiff base counterpart, and also the high probability of imine hydrolysis, formation of HNE-Schiff bases was neither expected nor detected. HNE modification of the targeted residues was confirmed using accurate mass data of the corresponding peptides as well as high-resolution MS/MS spectra collected from the HNE-treated samples.

All high resolution MS/MS spectra for HNE-modified peptides contained the previously reported (Roe *et al.*, 2007) diagnostic fragment ions at *m/z* 139.11 and 266.19 (in the case of histidine modification), representing the dehydrated protonated molecule of HNE and the immonium ion of HNE-modified histidine, respectively, as well as occasional neutral loss of HNE (156.1150 Da) (Figure 6.3). In searching for HNE-related (M+156.1) modifications, it is important to be able to distinguish between an Arg residue (156.1011 Da) and HNE addition (156.1150 Da), which can easily be achieved by high-resolution mass spectrometry. Figure 6.4 illustrates representative MS/MS spectra from the peptide containing His-48 with and without HNE modification. This example also depicts the case of the HNE adduct losing an additional molecule of water. All high-resolution MS/MS spectra for peptides characterized in their unmodified and HNE-modified forms are included in Appendix C.

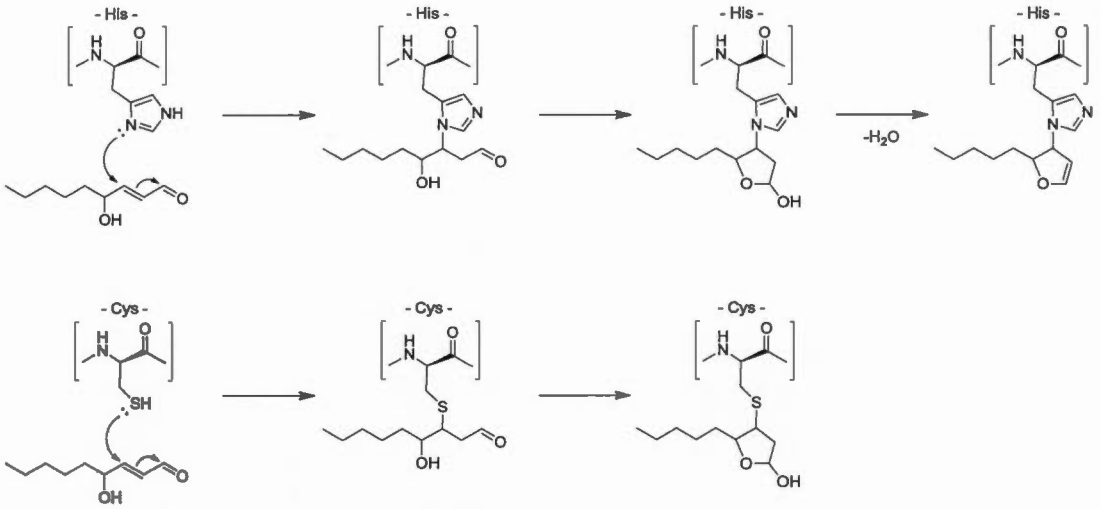


Figure 6.2 Mechanism of reaction for the nucleophilic addition of histidine (above) and cysteine (below) to HNE

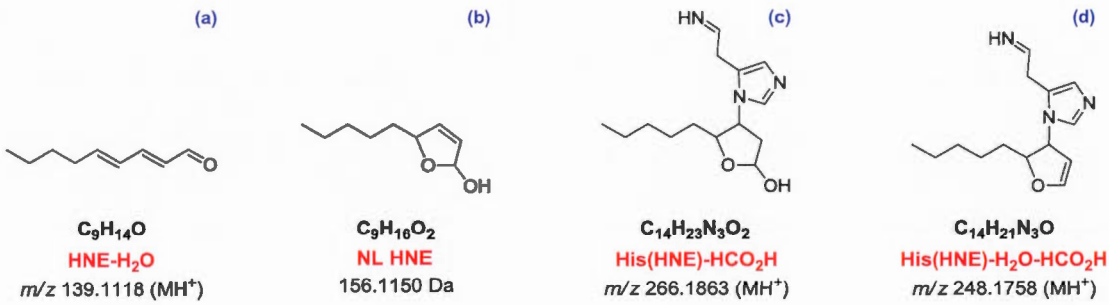


Figure 6.3 HNE adducts CID-generated diagnostic products corresponding to (a) cleaved dehydrated HNE, (b) neutral loss of HNE, (c) cleaved immonium ion of HNE-modified histidine, and (d) cleaved dehydrated immonium ion of HNE-modified histidine

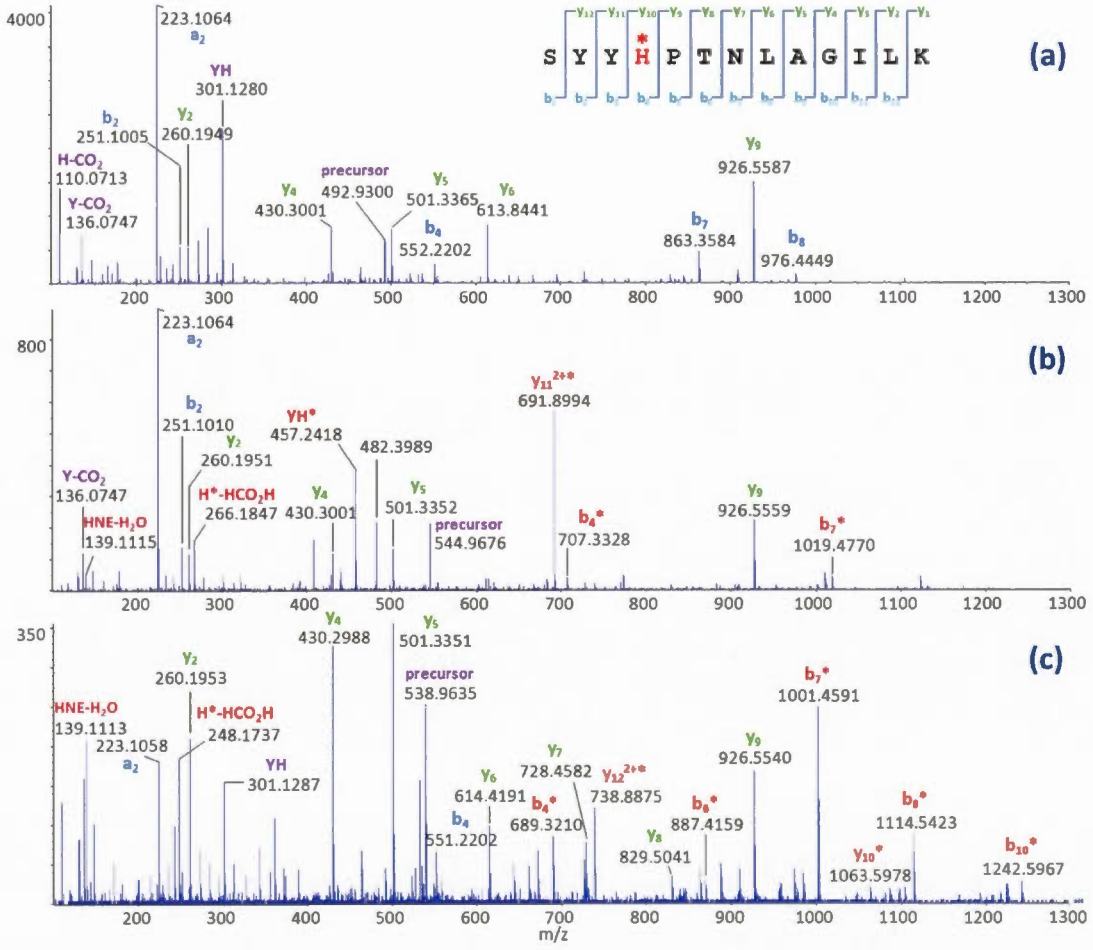


Figure 6.4 MS/MS spectra from the His-48-containing peptide (+3 charge state) in unmodified (a) and HNE-modified (b) forms. The *m/z* 139.11 and 266.18 are characteristic fragment ions associated with dehydrated protonated molecule of HNE and the immonium ion of HNE-modified histidine, respectively. In this example, a dehydrated modified peptide was also found, with a diagnostic ion at *m/z* 248.17 (c).

Table 6.1 lists the peptides found to contain HNE-modification sites. Among the identified HNE-modified peptides, three were non-tryptic (YYH^{48}PTN, LLAH^{187}AFPPGP and SH^{251}FMLPDDDVQG), while IDAAYEH^{340}PSH^{343}DLIFIFR represented the sole peptide containing two modification sites. Tryptic miscleavages were also detected in the case of TPDKC^{284}DPSLSLDAITSLR which could potentially be related to decreased enzyme efficiency due to the steric hindrance caused by the addition of HNE in the vicinity of the cleavage site. An example of multiple HNE additions on a single modification site was also observed on His-187, where doubly and triply HNE-modified peptides were detected at *m/z* 666.4023 (16.9 min) and *m/z* 744.4598 (21.5 min), respectively, due to a previously reported (Annangudi *et al.*, 2008) "domino-like multiple Michael adduction of HNE" to histidine. However, the singly modified peptide represented the most abundant species (in terms of MS sensitivity). All modified and unmodified peptides have been identified within 5 ppm mass accuracy.

Table 6.1 Distinct identified peptides encompassing HNE-modification sites (marked in red) within the MMP-13 structure

			Unmodified				Modified			
Residue	Distinct Peptide	z	theoretical *m/z*	measured *m/z*	error (ppm)	RT (min)	theoretical *m/z*	measured *m/z*	error (ppm)	RT (min)
His-48	SYYHPTNLAGILK	3	492.9330	492.9347	3.5	11.1	544.9714	544.9728	2.6	13.3
	YYHPTN	2	397.6771	397.6767	-0.9	2.7	475.7346	475.7348	0.5	9.1
Cys-96	CGVPDVGEYNVFPR	2	804.8774	804.8770	-0.4	12.1	854.4242	854.4245	0.4	14.8
His-187	LLAHAFPPGP	2	510.2873	510.2882	1.7	10.4	588.3448	588.3446	-0.3	13.7
His-251	SHFM(Ox)LPDDDVQG	2	688.7930	688.7907	-3.3	10.0	766.8505	766.8510	0.7	12.8
	SHFM(Ox)LPDDDVQGIQSLYGPGDEDPNPK	3	996.4503	996.4534	3.1	12.3	1048.4887	1048.4859	-2.6	13.9
Cys-284	TPDKCDPSLSLDAITSLR	3	663.6701	663.6720	2.8	12.4	696.7013	696.7037	3.4	14.9
His-312	LHPQQVDAELFLTK	3	546.9665	546.9683	3.2	11.4	599.0049	599.0053	0.7	13.2
His-340 His-343	IDAAYEHPSHDLIFIFR	4	511.7626	511.7650	4.7	13.0	589.8202	589.8226	4.1	16.9
His-387	ISAAVHFEDTGK	3	425.5507	425.5527	4.8	8.1	477.5891	477.5907	3.4	10.8

From a biochemical point of view, all the adducted amino acids play important roles in MMP-13 enzymatic function and are in regions of biological significance. More precisely, His-48 and Cys-96 are located on the pro-peptide (or activation peptide) which must be removed from the zymogen in order for the enzyme to be activated. Cys-96 is also a strictly conserved residue involved in the "cysteine switch" mechanism of activation (Van Wart and Birkedal-Hansen, 1990). Together with the three catalytic histidines (His-222, His-226 and His-232) and the "catalytic water" stabilized by Glu-223, Cys-96 forms a relatively stable complex with the catalytic zinc (Zinc 2), a structure known as the "metzincin-type zinc-binding consensus", preventing binding and cleavage of the substrate and is therefore essential for keeping the enzyme in its inactive form (Overall, 2000). HNE modification of this cysteine disrupts the zinc complex structure and could potentially lead to the unusual activation of MMP-13 in favor of type II collagen breakdown and therefore undesired cartilage degradation. His-251 is located within the specificity loop, giving MMP-13 a higher affinity towards type II collagen and is thought to be involved in an increased proteolytic activity if covalently modified (Aureli *et al.*, 2008). His-187 is one of the four residues (together with His-172, Asp-174 and His-200) extremely tightly bound to the structural zinc (Zinc 1) (Overall, 2000). All the other modification sites are part of the hemopexin-like domains, which play a key role in activation, inhibition and substrate-binding properties of all MMPs (Piccard *et al.*, 2007).

6.4.2 Reactivity of HNE Modification Sites

To compare the relative reactivity of the identified modification sites, rhMMP-13 (100 nM) was incubated with different HNE concentrations (0, 8, 40, 200, 1000 μM), and in a second test, with varying reaction times (1-120 min). The extent

of modification was measured by the ratio of modified-to-unmodified peak areas from high-resolution extracted ion chromatograms. These data are illustrated in Figure 6.5. From these experiments, exposure to HNE was found to have more of an effect on histidine adduction than cysteine modification. In particular, the modification of His-251 (in the specificity loop) as well as His-340 and His-343, located on a hemopexin-like domain, exhibited the highest dependence on both HNE concentration and reaction time. His-48, located in the pro-peptide region, is less concentration-dependent but is modified at high rates even after long incubation times. On the other hand, the extent of His-387 modification reaches a plateau after 30 minutes; however the reaction continues to increase at very large excess HNE. His-187 and His-312 both demonstrate less dependence on exposure to HNE than other histidines. The two cysteine modification sites (Cys-96 and Cys-284) were less affected by the reaction conditions tested, with Cys-284 being the least reactive of all found sites.

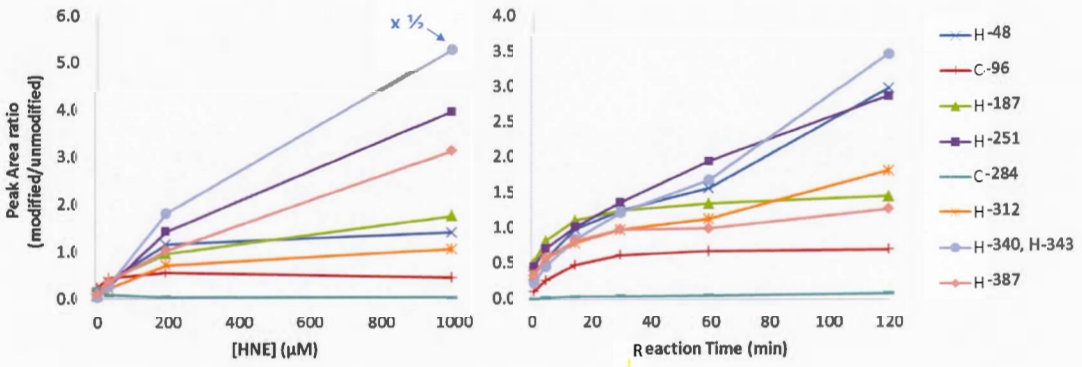


Figure 6.5 Graphic representation of modified/unmodified peak area ratios under different HNE concentrations (left) and incubation times at [HNE] = 200 μM (right)

6.4.3 Targeted MRM-Based Assay

Multiple-reaction monitoring (MRM) is a targeted technique enabling the selective detection of specific peptides with enhanced sensitivity, therefore is useful for monitoring specific peptides of interest in complex samples. An MRM-based method was developed for directly monitoring each of the peptides which were previously found to be modified by HNE, in both unmodified and modified forms. This enables higher sensitivity for confirming the presence of modified peptides from complex samples. Collision energies were optimized for the most sensitive MRM transitions (most abundant product ions for each peptide). Detailed information on the optimized MRM method can be found in Table 6.2. This targeted method was useful for the analysis of biological samples due to decreased interfering signals present in complex matrices.

6.4.4 Analysis of Cell Culture Samples

Complexity of biological samples is one of the major current limitations of proteomic investigations affecting both the quality and quantity of the MS-acquired data (Chandramouli and Qian, 2009). However, recent advancements have improved protein identification for even the least abundant proteins in biological samples. Modern fractionation techniques, high resolution LC separations, sophisticated search algorithms, and most importantly, the advancement of high-resolution and high-speed MS instruments have enabled in-depth analysis of the proteome (Mann *et al.*, 2013).

Table 6.2 Optimized *m/z* values for the precursor and product ions and collision energies (in parentheses) for MRM analysis of unmodified and HNE-modified peptides detected for MMP-13

Distinct Peptide		Precursor ion *m/z*	Product ion *m/z* (CE) 1	2	3
SYYH[48]PTNLAGILK	unmod	492.9	430.3 (22)	501.3 (22)	926.6 (20)
	mod	545.0	139.1 (25)	260.2 (32)	266.2 (38)
YYH[48]PTN	unmod	397.7	235.1 (26)	331.2 (23)	464.2 (16)
	mod	475.7	139.1 (22)	235.1 (32)	266.2 (32)
C[96]GVPDVGEYNVFPR	unmod	804.9	317.1 (39)	419.2 (37)	981.5 (35)
	mod	854.4	139.1 (32)	419.2 (43)	981.5 (37)
LLAH[187]AFPPGP	unmod	510.3	367.2 (19)	506.3 (29)	653.4 (17)
	mod	588.3	266.2 (41)	367.2 (25)	662.4 (35)
SH[251]FM(Ox)LPDDDVQG	unmod	688.8	519.2 (33)	632.3 (29)	1074.4 (29)
	mod	766.8	139.1 (34)	675.3 (27)	788.4 (34)
SH[251]FM(Ox)LPDDDVQGIQSLYGPGDEDPNPK	unmod	996.4	968.4 (39)	1025.4 (43)	1188.5 (39)
	mod	1048.5	968.4 (41)	1025.4 (43)	1188.5 (39)
TPDKC[284]DPSLSLDAITSLR	unmod	663.7	476.3 (28)	775.4 (30)	975.5 (32)
	mod	696.7	139.1 (35)	476.3 (30)	775.4 (33)
LH[312]PQQVDAELFLTK	unmod	547.0	251.1 (36)	476.3 (25)	604.3 (20)
	mod	599.0	139.1 (28)	266.2 (38)	407.3 (36)
IDAAYEH[340]PSH[343]DLIFIFR	unmod	511.8	435.3 (25)	582.3 (23)	695.4 (19)
	mod	589.8	139.1 (29)	266.2 (43)	435.3 (35)
ISAAVH[387]FEDTGK	unmod	425.6	305.2 (23)	549.2 (27)	833.4 (20)
	mod	477.6	139.1 (30)	266.2 (32)	305.2 (25)

In the present study, three different workflows were employed to enrich the culture media extracted from chondrocytes of OA patients: molecular weight cut-off (MWCO) filtration, protein pre-fractionation and immunoprecipitation. The LC-MS/MS analysis of the MWCO sample (1.5 ml, 2.5 μg/ml) filtered to include only the molecules larger than 30 kDa led to the identification of 24 proteins; however, MMP-13 was not present in this list. By extracting accurate masses from the list of target peptides, a few unmodified peptides were detected as well as a prominent signal for the HNE-modified peptide containing His-48 ($YYH^{48}PTN$, *m/z* 475.7339 at 8.2 min).

In a second experiment, the MWCO-filtered sample was separated into 8 fractions using mixed-bed weak cation/anion exchange (CATWAX) to further reduce complexity. Mixed-bed and dual-column cation/anion exchange fractionations have been previously reported (Zhang *et al.*, 2012; Havugimana *et al.*, 2007; Motoyama *et al.*, 2007) as efficient protein separation methods. Using ProteinPilot, 110 proteins were found, with the majority concentrated in fractions 5-6 (Table 6.3). The average sequence coverage of the identified proteins was 13.6% with an average of 5 confident peptides per protein. Although MMP-1, MMP-2 and MMP-3 were detected, the ProteinPilot search did not result in the identification of MMP-13.

Immunoprecipitation (IP) is an efficient and highly specific technique to isolate and concentrate a particular protein from a complex biological matrix (Wilson and Walker, 2010). When IP was performed on the cell culture sample, LC-MS/MS analysis led to improved results with a total number of 6 identified proteins (as well of a few contaminating keratins) with MMP-13 having the highest protein coverage (45.9%), and the formation of HNE/MMP-13 adduct was confirmed by Western blot using an anti-HNE antibody (Figure 6.6). Similar to the results from CATWAX fractions, MMP-1 and MMP-3 were also detected. A list of proteins found in the IP-enriched sample is given in Table 6.4.

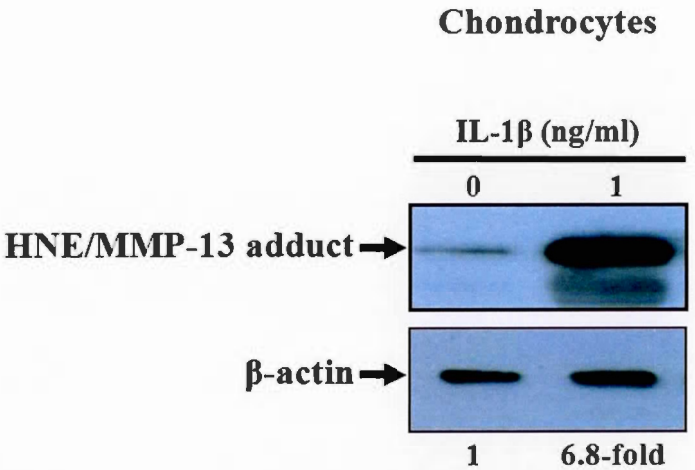


Figure 6.6 Evidence of HNE binding to MMP-13 by immunoprecipitation. 100 mg of total protein were immunoprecipitated using anti-MMP-13 antibody and then subjected to Western blotting using anti-HNE antibody. Total proteins from IL-1β-treated chondrocytes showed a 6.8-fold increase in HNE/MMP-13 adduct levels as compared to untreated cells. β-Actin was used as a loading control in Western blot analysis.

Table 6.3 Distribution of the identified proteins and peptides in the CATWAX fractions collected from human chondrocyte cultures

Fraction	1	2	3	4	5	6	7	8	1-8 combined
# Proteins (1% FDR)	10	30	0	0	47	58	24	8	110
# Peptides (1% FDR)	45	123	41	50	232	319	69	22	615

Table 6.4 List of the identified proteins (1% FDR) in human chondrocyte cultures enriched by immunoprecipitation

Representative Accession	Name	Confident Peptides	Sequence Coverage (%)
sp|P45452|MMP13_HUMAN	Collagenase 3 (MMP-13)	39	45.9
sp|P08254|MMP3_HUMAN	Stromelysin 1 (MMP-3)	27	24.7
sp|P51884|LUM_HUMAN	Lumican	4	12.7
sp|P07093|GDN_HUMAN	Glia-derived nexin	1	3.5
sp|P03956|MMP1_HUMAN	Interstitial collagenase (MMP-1)	1	2.8
sp|Q5VTE0|EF1A3_HUMAN	Putative elongation factor 1-α-like 3	1	2.4

A targeted analysis was performed on the IP-enriched sample to screen peptides encompassing all possible HNE modification sites as previously detected during *in vitro* experiments. All the modification sites were detected in the IP sample and the corresponding peptides were confirmed using accurate mass MS and MS/MS spectra on the QqTOF system as well as LC-MRM. Each of the modified peptides found represent unique sequences from MMP-13, except for LLAH^{187}AFPPGP, which could also originate from MMP-9, however the latter protein was not detected in the biological sample. Interestingly, the peaks related to the modified peptides containing His-340, His-343 and His-387 (IDAAYEH^{340}PSH^{343}DLIFIFR and ISAAVH^{387}FEDTGK), all located on the hemopexin-like domains, were more prominent in the culture sample (relative to other modification sites) than in the HNE-treated recombinant MMP-13 samples. This increased preference can be regarded as an indication that these sites may be more accessible than other identified modification sites in a cellular context. On the other hand, cysteine modifications were much less evident in the cell culture sample.

The modification at Cys-96 was confirmed by accurate mass and MRM analysis, however, was less prominent than in the *in vitro* sample. No peptide signal was found containing Cys-284, in its modified or unmodified forms. Therefore, sufficient evidence was not found to support the "cysteine-switch" mechanism (Morquette *et al.*, 2006) as the principal cause for the increased activity of MMP-13 in OA patients. The presence of each of the modified peptides in the biological sample was confirmed using high-resolution TOF-MS (accurate mass of precursor ion and associated charge state confirmation), full scan HR-MS/MS, and MRM analysis. All modified peptides were confirmed by accurate precursor mass (and RT matching) as well as MRM signal, and many also had enough intensity to produce high quality HR-MS/MS spectra (Table 6.5). Figure 6.7 illustrates XICs from the HNE-modified peptides from the treated recombinant sample and from the immunoprecipitated cell culture sample.

Table 6.5 LC-MS/MS data obtained from the analysis of human chondrocyte cultures to confirm HNE-modification sites of MMP-13

Residue	Accurate Mass	MRM	MS/MS	Confident Peptide
His-48	√	√	√	**YYHPTN**
Cys-96	√	√	x	**CGVPDVGEYNVFPR**
His-187	√	√	x	**LLAHAFPPGP**
His-251	√	√	√	**SHFM(Ox)LPDDDVQGIQSLYGPGDEDPNPK**
His-312	√	√	√	**LHPQQVDAELFLTK**
His-340	√	√	x	**IDAAYEHPSHDLIFIFR**
His-343	√	√	x	**IDAAYEHPSHDLIFIFR**
His-387	√	√	√	**ISAAVHFEDTGK**

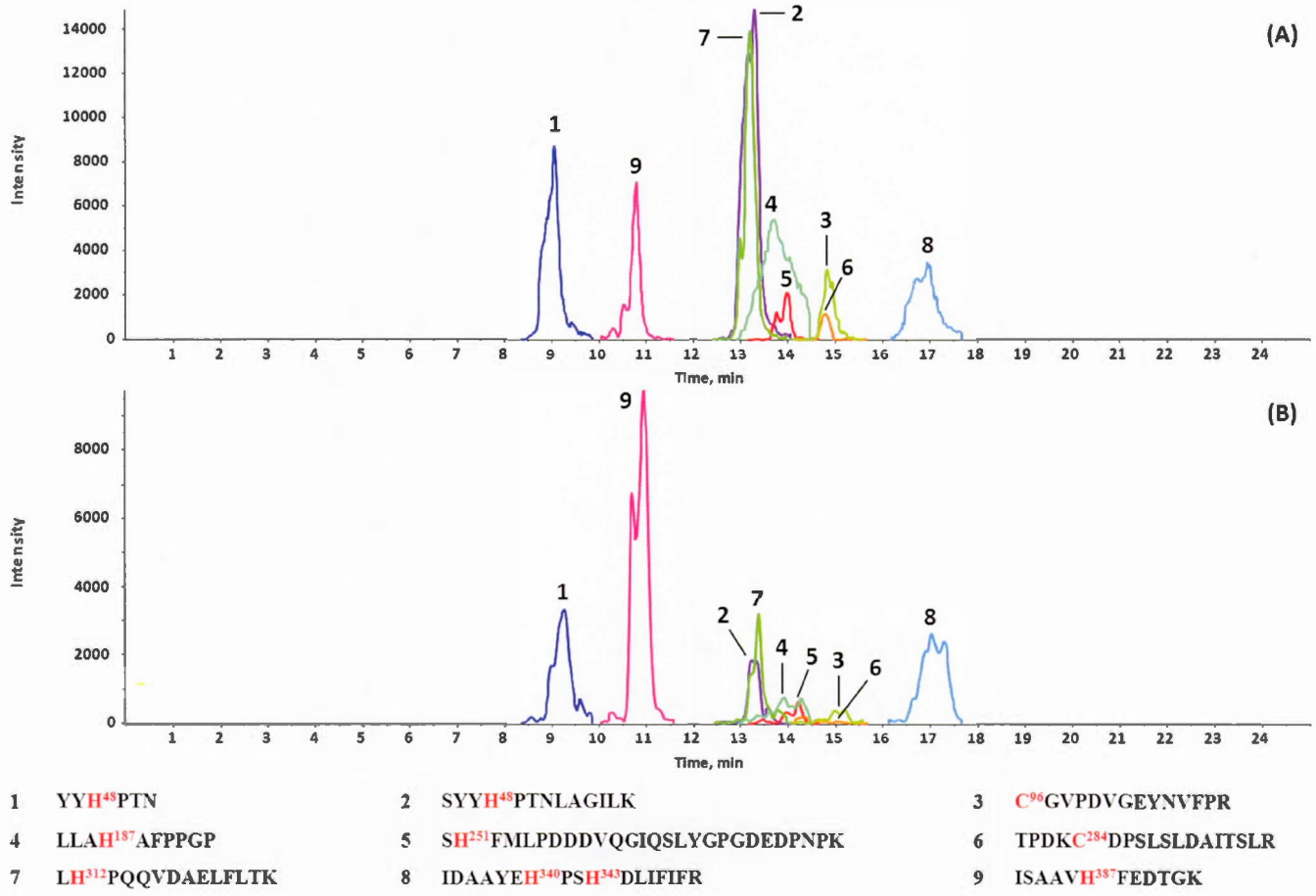


Figure 6.7 Overlaid extracted ion chromatograms of digested peptides containing HNE-modifications obtained from HNE-treated rhMMP-13 (A) and IP-purified MMP-13 from cultured chondrocytes (B)

6.5 Conclusions

In conclusion, covalent modification of MMP-13 by HNE was investigated using a LC-HR-MS/MS method developed to pinpoint and characterize HNE modification sites combined with a MRM-based analysis for targeted monitoring of these modification sites. Incubation of MMP-13 with HNE *in vitro* demonstrated that seven histidine and two cysteine residues were covalently modified in the protein sequence. Time-course and concentration-dependent experiments revealed that histidine adduction was more dependent on HNE concentration and reaction time than cysteine modification. Since previous data had suggested that certain covalent modifications may lead to activation or over-expression of the enzyme resulting in accelerated progress of joint degenerative diseases such as OA, the developed method was employed to screen HNE modifications in human chondrocyte cultures extracted from OA patients. The results of this study elucidated that the HNE modifications previously identified *in vitro* were also present in cell culture samples. Three histidine modification sites (His-340, His-343 and His-387) seemed to be more extensively targeted by HNE in cells from OA patients compared to the relative contribution of each site obtained from *in vitro* incubations.

The main goal of this study was to develop an analytical method to characterize covalent modification of MMP-13 by HNE. The results from the OA patients were regarded as a proof-of-concept to demonstrate the viability of the developed method, although further investigations could be performed on biological samples using the developed methodology, including comparing severity of disease with negative controls and to probe the importance of HNE modification under different cell treatment conditions. These follow-up studies have the potential to increase our understanding of the development of chronic diseases related to MMP-13 activity and may lead to novel treatment avenues.

6.6 Acknowledgments

M. G. would like to thank GRASP (*Groupe de Recherche Axé sur la Structure des Protéines*, FRSQ) and *la Fondation de l'UQÀM* (*Bourses d'Excellence des Cycles Supérieurs*, FARE) for financial support.

6.7 Supporting Information

High-resolution MS/MS spectra of the unmodified and HNE-modified peptides associated with each MMP-13 modification site characterized in *in vitro* sample, demonstrating the formation of the HNE/MMP13 adduct (Appendix C).

CHAPTER SEVEN

SUMMARY AND CONCLUSIONS

Throughout the previous chapters, efforts were made to demonstrate the importance of each step of the analytical method development for effective detection of protein covalent modifications in complex biological samples using LC-MS/MS. Two well-studied reactive metabolites, NAPQI and HNE, were selected and used as models to establish a systematic approach for the identification of protein adduction sites *in vitro*. To achieve this, three following questions had to be answered:

1. What are the optimal sample preparation conditions to minimize possible interferences and maximize the detection potential of low-abundant covalently modified peptides in complex mixtures?

2. What is the most efficient combination of separation techniques to reduce sample complexity for the most efficient LC-MS/MS analysis?

3. How to process large quantities of LC-MS/MS data to filter interferences, eliminate false positives, detect and verify covalently modified peptides in xenobiotic-treated samples.

The present dissertation attempted, in a methodical way, to answer the above questions based on the results obtained through each part of the optimization. The following sections summarize these findings and present a global outlook of the steps to take in the future to further ameliorate the established methods, test their efficiency for other research designs, and clinically validate the current results.

7.1 Summary of Findings

The focus of Phase I of this research was to answer the first question for finding the efficient preparation of liver microsomal samples, planned to be employed as a relevant model to study protein covalent modification by reactive metabolites. In order to answer this question, three experimental procedures needed to be established based on the existing knowledge on the analysis of liver microsomes (Qiu *et al.*, 1998; Rombach and Hanzlik, 1999; Evans *et al.*, 2004; Plant, 2004; Shin *et al.*, 2007; Liebler, 2008; Tzouros and Pähler, 2009; Kalgutkar, 2011; Yang *et al.*, 2014): (1) protein solubilization, (2) digestion, and (3) the initial protein identification approach using LC-MS/MS.

To maximize the solubility of microsomal proteins, reagents from different classes of solubilizing agents were tested: SDS, FC-14 and SDC (ionic detergents), DDM (non-ionic detergent), guanidinium hydrochloride (chaotrope), methanol, formic acid and TFA (organic solvents and carboxylic acids). SDS demonstrated the best performance in terms of number of proteins and peptides, and average protein sequence coverage. To optimize digestion conditions, trypsin, pepsin, endoproteinase Glu-C, and cyanogen bromide were compared separately, and in combination. Trypsin produced better results particularly for hydrophilic proteins, whereas pepsin yielded higher sequence coverage for membrane proteins. Sequential trypsin-pepsin dual digestion did not work as efficiently as each enzyme separately, and combining the two digestion steps in parallel was better only if samples were digested and fractionated separately and LC-MS/MS data were then co-processed at the protein search level. To compare the different treatment strategies, a data analysis workflow was also developed using ProteinPilot protein search software and free access web-based bioinformatics tools. Figure 7.1 depicts the steps involved in this workflow. Using these methods, a total of 1095 proteins were identified in rat liver microsomes at 1% FDR, out of which 213 proteins (19.5%) were IMPs.

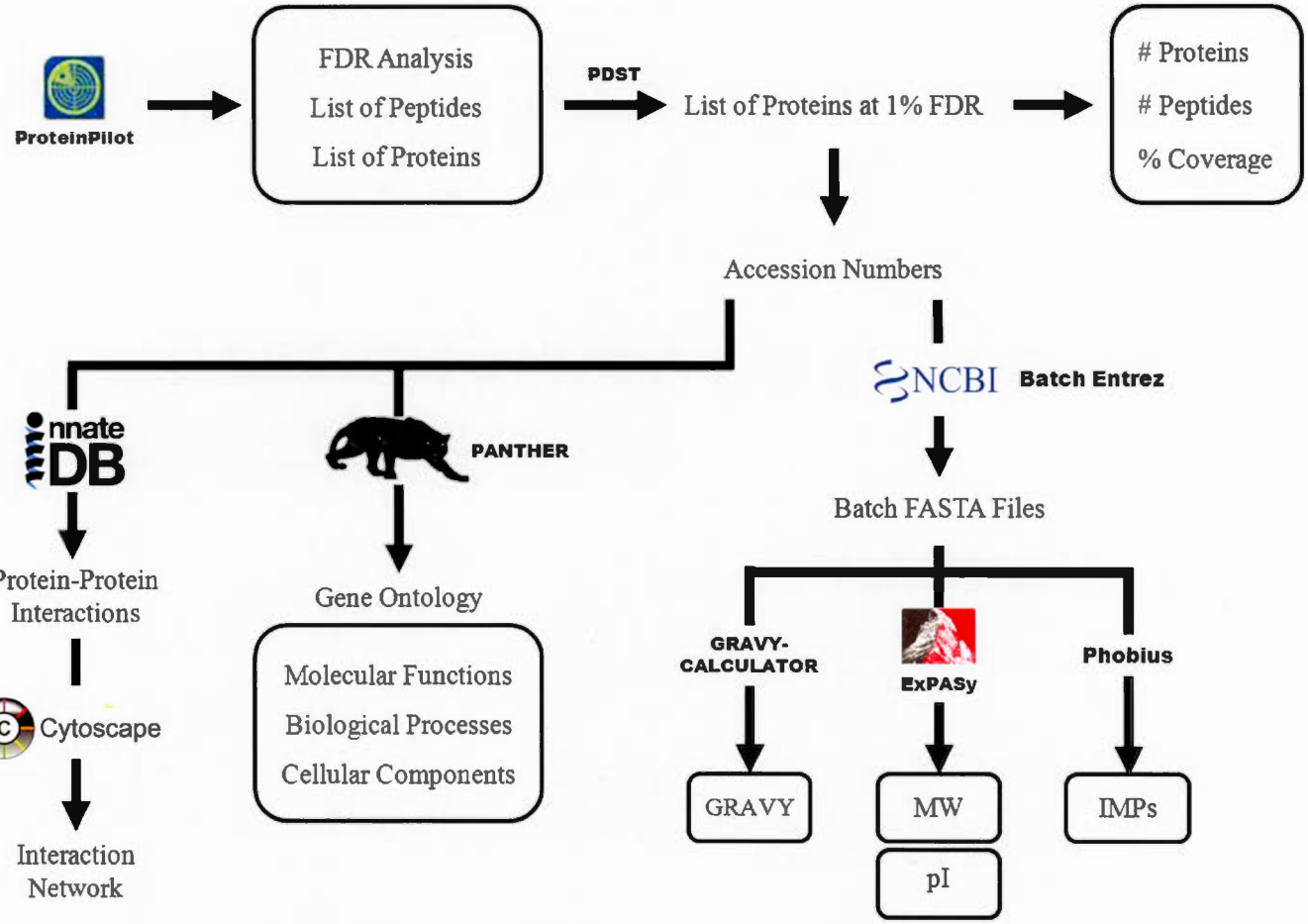


Figure 7.1 Workflow for the bioinformatics analysis of rat liver microsomal proteins. FDR: false discovery rate; GRAVY: grand average of hydropathy; IMP: integral membrane protein; MW: molecular weight; PDST: ProteinPilot™ descriptive statistics template.

The methods developed in Phase I were used to treat samples and data in Phases II and III. The publication compiling the results of this work (Golizeh and Sleno, 2013) has since been cited by other studies aiming at improving proteomics analysis performance using multi-enzyme digestion strategies (Lin *et al.*, 2013; Lin *et al.*, 2014; Zhang, 2015; Dunston *et al.*, 2015; Nardiello *et al.*, 2015).

Four combinations of ion-exchange fractionation with RP-LC were examined in Phase II to optimize peptide separation prior to the MS/MS analysis. Different modes of ion exchange chromatography were compared, separately and in tandem, for protein fractionation and then tested against SCX peptide fractionation in 2D and 3D configurations. The 2D approach combining SCX with RP-LC yielded the best results. The four multidimensional LC-MS/MS approaches compared in this study are demonstrated in Figure 7.2.

The core–shell RP column used in this study enabled higher resolving power and enhanced peak capacity. Core–shell particles are composed of a solid silica core and a thin porous outer layer, and are used in LC columns employed in the analysis of complex mixtures. In theory, a core–shell particle's diffusion behaviour resembles a fully porous counterpart with a diameter equal to the thickness of the core–shell particle's outer layer. These packing materials therefore enable UHPLC separations with virtually smaller particles without a drastic increase in backpressure due to reduced particle size according to Darcy's law (De Vos *et al.*, 2012). Furthermore, since it is only the outer layer that acts as the stationary phase, columns filled with core–shell particles have less solvent capacity thus enabling a faster elution compared to the conventional columns.

The RP-LC separation was optimized in Phases I and II. However, based on initial tests it was decided to use a longer gradient in Phase III to allow for a more efficient separation of APAP-modified peptides from the unmodified counterparts.

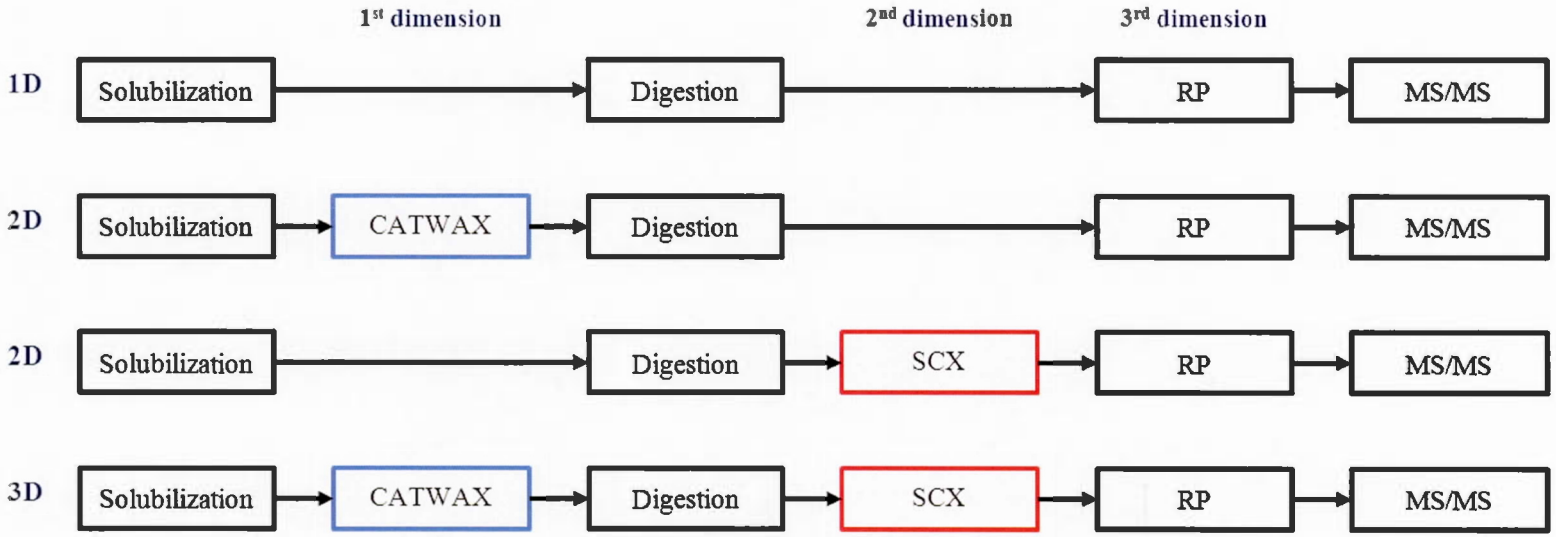


Figure 7.2 MD-LC approaches employed to maximize peptide separation efficiency for proteomics analysis of rat liver microsomes and S9 fractions (in Phase II). CATWAX: mixed-mode weak cation/anion exchange; RP: reversed-phase chromatography; SCX: strong cation exchange

MS/MS data were acquired in this study using information-dependent (IDA) (or data-dependent, DDA) acquisition. In this approach, precursor ions are "selected" from a MS ("survey") scan based on a set of given selection criteria, activated, and subjected to the MS/MS ("dependent") scan (Figure 7.3). However, since a portion of precursor ions will not be selected for MS/MS, usually due to low signal intensity, some information, particularly related to less-abundant peptides, can get lost in this type of analysis. Nevertheless, DDA is an efficient way to increase data acquisition speed and can generate more MS data with higher quality from LC peaks, compared to the conventional LC-MS/MS methods. DDA parameters including the maximum number and threshold signal intensity of selected precursor ions, MS/MS acquisition time, and dynamic exclusion conditions were optimized in Phase II leading to a 46% increase in the number of proteins (1% FDR), 20% more average protein sequence coverage, 48% more IMPs and 19% shorter cycle time compared to Phase I. DDA also allows for selective acquisition of MS/MS spectra for a given list of precursor ions ("inclusion list") enabling better characterization of complex mixtures. This strategy was used to confirm the structure of modified peptides in Chapters 5 and 6.

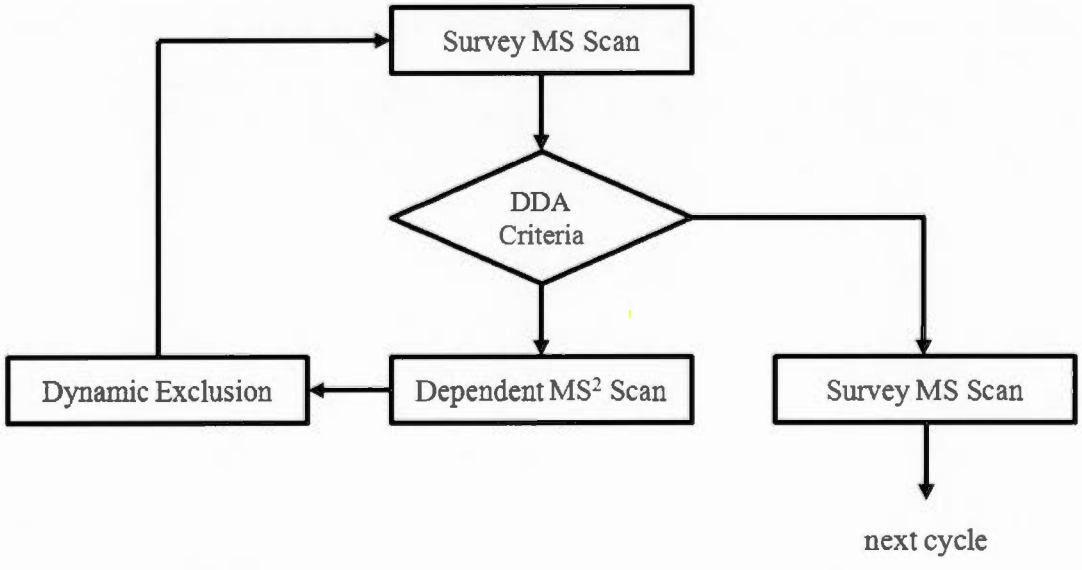


Figure 7.3 A typical data-dependent acquisition (DDA) cycle with dynamic exclusion

Once the different steps of the analytical method were optimized, liver S9 and microsomal fractions from rat, mouse and human were analyzed for an inter-species comparison. GO analysis demonstrated that most identified proteins had catalytic, receptor or binding properties and were primarily involved in cellular and metabolic processes and response to the stimuli. This information provides insights into what types of biological processes can potentially be affected by covalent modification of these proteins, which can be useful in molecular toxicology and pharmacology studies (Golizeh *et al.*, 2015c).

In Phase III, an approach was developed for the identification of covalently modified proteins in rat liver microsomes incubated with APAP. A large focus was put to devise a data processing workflow to effectively detect APAP-modified peptides with a high capacity for removing potential false positives. Peptide spectral matching, with an increased APAP-modification probability, was combined with differential analysis of LC-MS signal intensities in treated and control samples, APAP-Cys adduct-related CID-product ion screening, and searching for peak-pairs with an APAP-adduction mass difference. Putative modified peptides found by more than one data analysis approach were subjected to multiple levels of data verification and five APAP-modified peptides were finally confirmed by exact mass and MS/MS spectra highlighting covalent modification of four proteins: MGST1, CP2C6, NU5M and MADD. These proteins are believed to be associated with oxidative stress and their covalent modification could potentially be linked to the cellular damage involved in APAP-induced toxicity (Golizeh *et al.*, 2015a).

Figure 7.4 illustrates the overall scheme of the approach developed throughout Chapters 2 to 5 of this dissertation for identification of reactive metabolite protein targets using LC-MS/MS.

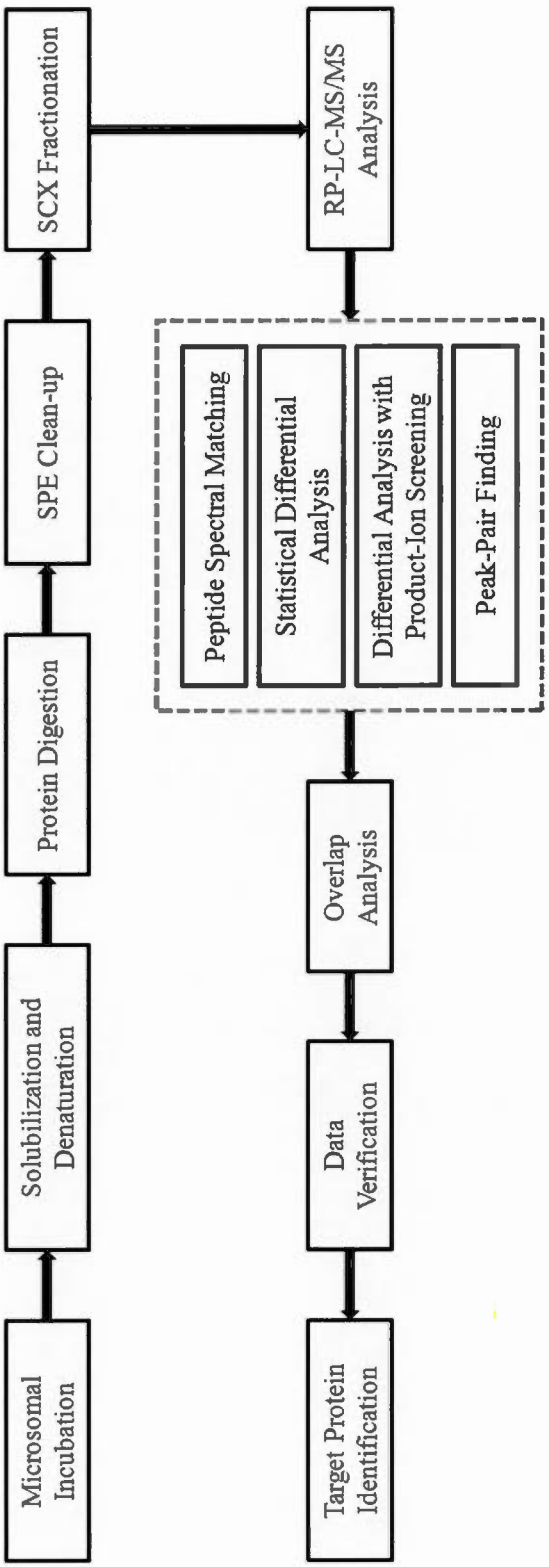


Figure 7.4 The analytical approach developed throughout this dissertation for identification of reactive metabolite protein targets using LC-MS/MS

In Chapter 6, a more targeted method was developed to examine covalent modification of MMP-13 by HNE *in vitro* using LC-HR-MS/MS and LC-MRM. Analysis of HNE-incubated rhMMP-13 demonstrated that seven histidine and two cysteine residues were HNE-modified. Kinetic and concentration-dependent studies allowed for the comparison of the reactivity of these modification sites. The same HNE-modifications were detected on MMP-13 in chondrocyte culture extracts from osteoarthritis patients, providing insights to support the previously suggested (Morquette *et al.*, 2006) potential involvement of HNE and oxidative stress in the development of OA. The publication describing these results (Golizeh *et al.*, 2014) has been cited by other studies on detection of protein lipoxidation products (Aldini *et al.*, 2015), as well as qualitative (Milic *et al.*, 2015) and quantitative (Yang *et al.*, 2015) analysis of protein alkylation by HNE.

In summary, methods were developed using LC-MS/MS for the identification of reactive metabolite protein targets in complex mixtures, determination of their modification sites, and screening under different conditions. These methods can potentially be adapted for the analysis of similar sample types from animal models and human samples to provide useful information on protein adduction in both *in vitro* and *in vivo* studies.

7.2 Limitations and Perspectives

The study of reactive metabolites is a dynamic and progressive area of research with increased interest over the past few decades. Although initial studies were mainly focused on structural characterization of reactive metabolites (Kalgutkar, 2011), suggested links between their reactions with macromolecules and the resulting toxicity has drawn more attention to covalent binding reactions and identification of

their potential targets, an area which gradually became more feasible by means of technological advancements in bioanalysis.

In this work, four proteins were systematically identified to be modified by APAP in rat liver microsomes *in vitro* using a 2D-LC-MS/MS approach. The scope of this study is therefore limited to: (1) APAP protein targets, (2) rat liver microsomes, (3) covalent binding occurring *in vitro*. From a technical point of view, the method involved an unbiased separation of modified and unmodified peptides, and did not incorporate targeted enrichment of APAP-modified sequences prior to the MS/MS analysis. This work has paved the road for further improvements in scope, flexibility and efficiency of analysis for other types of molecules and samples. Some of these improvements and suggestions for future research will be discussed in the next few pages.

The method optimized in the preliminary experiments was ultimately utilized to identify APAP-modified proteins in a microsomal sample. However, with slight modifications, this method can be virtually used for any other reactive metabolite. For example, HNE is the most studied lipid-derived electrophile and one of the most abundant, reactive, and toxic endogenous metabolites (Yang *et al.*, 2015). However, unlike APAP or other drugs, HNE is already a reactive species and does not require biotransformation prior to the 2D-LC-MS/MS analysis. Several studies have previously suggested methods for characterization of protein-HNE adducts (for a recent review see Aldini *et al.*, 2015).

A preliminary experiment was conducted to study HNE covalent binding to liver microsomal proteins in parallel to the APAP-protein adducts analysis described in Chapter 5. Initial data processing demonstrated that, compared to APAP, a remarkably larger number of proteins were targeted by HNE. Future work would employ a comprehensive data analysis workflow, as developed for APAP, to determine rat liver microsomal HNE targets *in vitro*. The overlap between APAP and

HNE targets can also be interesting as it reveals what proteins are more susceptible to electrophilic attacks potentially opening doors to new therapeutic or protective approaches against reactive metabolite-mediated toxicity.

Another perspective from the work detailed in this thesis is the application of the optimized method to *in vivo* samples. As discussed in Chapter 6, HNE-modification sites found through *in vitro* incubations with MMP-13 were also detected from cell culture extracts, indicating that the reactivity of these sites towards HNE does not limit to simple protein incubations. Similar experiments can be designed using animal models to test if the identified APAP target proteins can also be found in liver homogenates of animals receiving a toxic APAP dose or patients diagnosed with APAP-induced toxicity. In this regard, several groups of rats and mice have been dosed with APAP. Livers were removed after different time points, and homogenized with solubilization using SDS and urea/thiourea. The same 2D-LC-MS/MS procedure has been employed as for the RLM samples. These results, once compiled, will be able to reveal if any of the previously detected APAP-adducted proteins were also found to be modified *in vivo*. In addition, these results can help identify other potential APAP targets not found through the RLM *in vitro* experiment. Moreover, the study of APAP-modifications at different time-points and dosing levels can provide insights into relative reactivity of the target proteins towards APAP in rat and mouse. These findings and the eventual results obtained from human studies could shed light on the mechanisms proposed for APAP-induced hepatotoxicity (Jaeschke *et al.*, 2013; Jiang *et al.*, 2015).

The approach developed in this thesis was used to identify protein adduction without any bias in the purification of modified and unmodified peptides. Although multi-dimensional separation and data-dependent acquisition greatly reduce the spectral complexity allowing for efficient MS/MS detection, less-abundant modified peptides may still be masked by the large excess of unmodified counterparts present

in the sample (Tzouros and Pähler, 2009). To address this challenge, more targeted methods have been developed using affinity tagging to selectively fish out the adducted peptides prior to the LC-MS/MS analysis. Affinity enrichment techniques have recently provided new opportunities to increase the specificity and selectivity of protein adduct detection methods. Biotin-tagging was first introduced by Adam *et al.* (2002) as an enrichment strategy to improve ICAT quantitative proteomics. Shin *et al.* (2007) developed two biotin-coupled reagents incorporating iodoacetamide and maleimide moieties to selectively bind to the thiol groups of cysteine-containing peptides (Figure 7.5 a–b).

Another approach, known as MeCAT,[13] was introduced by Ahrends *et al.* (2007) to enable selective separation and detection of cysteine-containing peptides tagged by a metal-chelating group (DOTA[14]) with enhanced sensitivity using ICP-MS (Figure 7.5 c). A similar biotin-tagging approach has been investigated in our group by LeBlanc *et al.* (2015) for selective enrichment of APAP-modified peptides using alkyne analogs of APAP (APAPyne) with "click chemistry". Using this approach, adducted peptides can be conjugated with an azide-containing biotinylation reagent which allows for subsequent enrichment of peptides via streptavidin affinity separation (Figure 7.6). A caveat of this type of method of course is the specific tagging of the compound to be studied, and therefore it is not applicable to all drugs or reactive species.

---

[13] Metal-Coded Affinity Tagging
[14] 1,4,7,10-Tetraazacyclododecane-1,4,7,10-tetraacetic acid

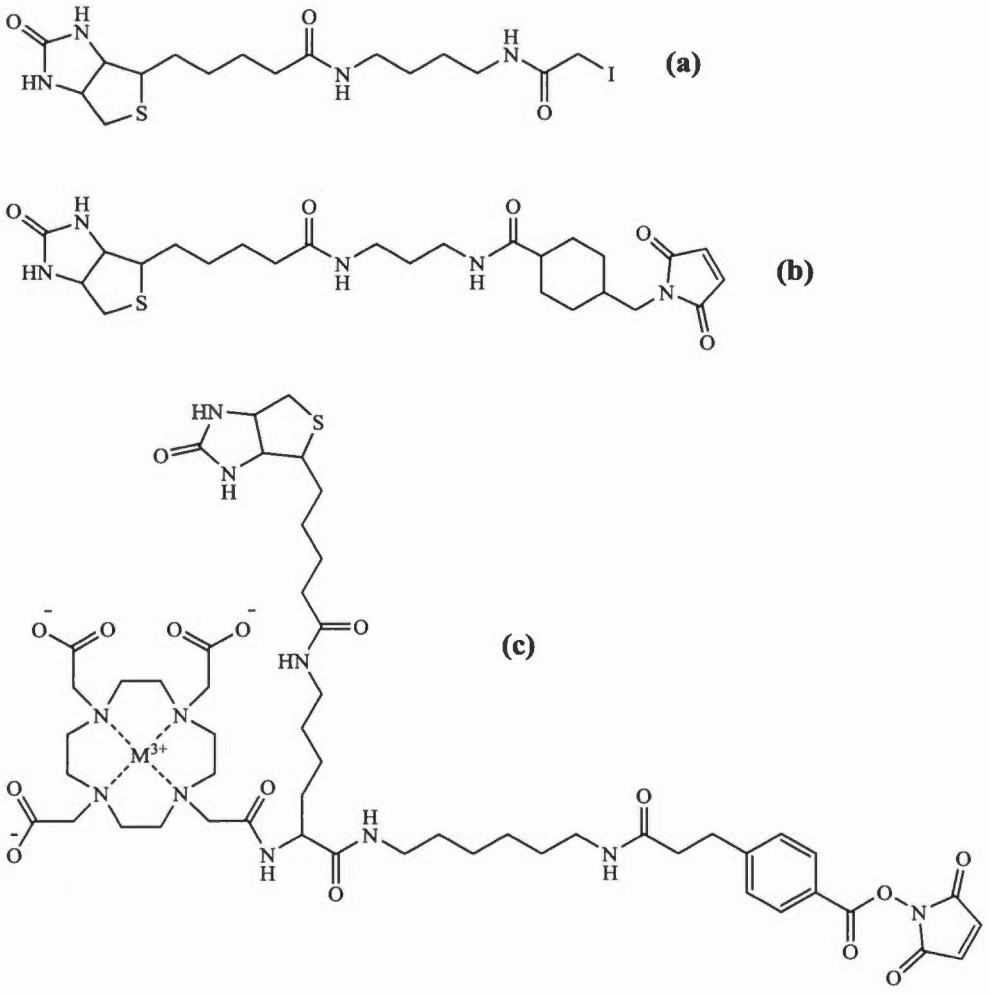


Figure 7.5 Biotin-coupled reagents for selective enrichment of peptides using affinity separation incorporating iodoacetamide (**a**) and maleimide (**b** and **c**) moieties for specific conjugation with the thiol functionality of cysteine-containing peptides

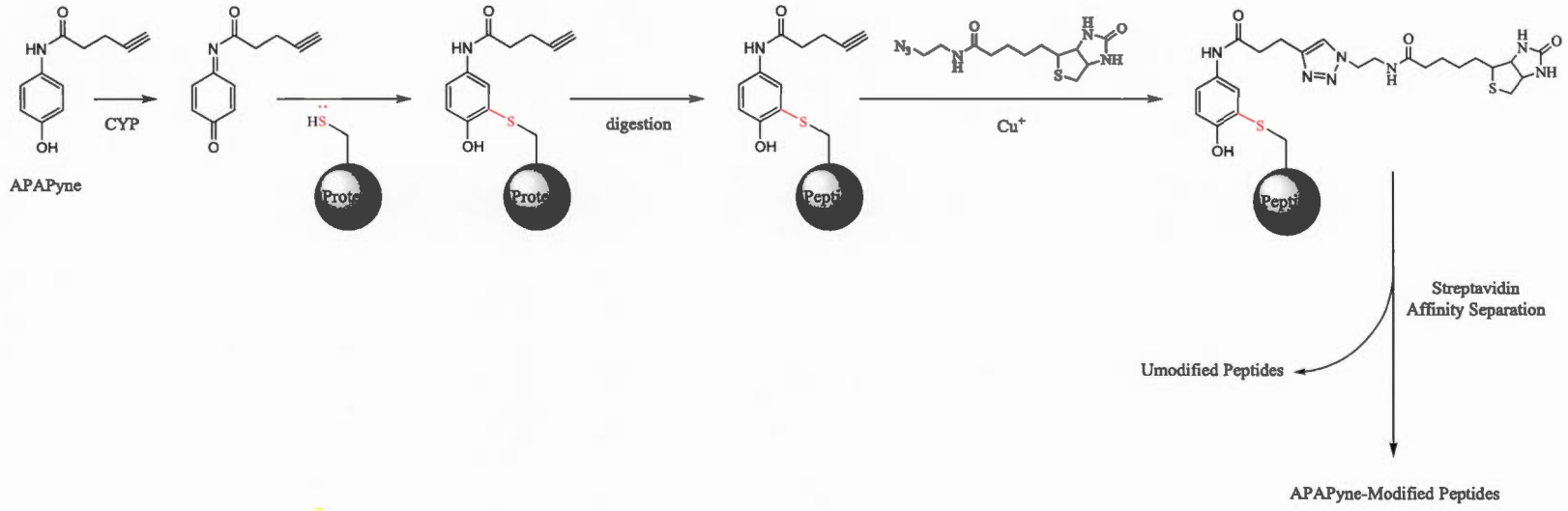


Figure 7.6 Sample preparation strategy for affinity enrichment of APAPyne-modified peptides using click chemistry

Another strategy to improve the separation capacity, detection sensitivity, and analysis throughput of the present method could be the application of micro- or nano-HPLC. This can be achieved by using reduced-dimension or capillary LC columns packed with sub-micron-sized (< 2 μm) C_{18}-bonded porous particles. The major advantage of using such columns is their ability to work with small sample volumes at low flow rates decreasing band dilution on the column hence offering enhanced sensitivity compared to conventional LC. Other benefits of micro/nano-LC are reduced solvent consumption, easy temperature control due to the fast and effective heat transfer on small columns, and high separation efficiency and peak capacity due to decreased sorbent particle diameter. However, since the pressure generated across a LC column is proportional to the square of the sorbent particle diameter, micro/nano-LC columns require pumps that can ensure high backpressure. Moreover, since peak volume decreases dramatically with column dimensions, elution band broadening can considerably deteriorate the separation if extra-column void volume (capillary connections, detector cell volume) is not minimized. This of course demands fast data acquisition using mass analyzers with high scanning rates (Šesták *et al.*, 2015). Microfluidic chip-based nano-LC has effectively dealt with these technical challenges and shown in practice to be more efficient than capillary LC columns (Liu *et al.*, 2013). In the HNE-MMP13 adduct detection experiment discussed in Chapter 6, a signal-to-noise ratio improvement of up to 40 times was obtained using a micro-LC (Eksigent Micro-LC 200) system on a Halo C_{18} 0.5 x 150 mm column packed with 0.5 μm porous silica-layer particles at 20 μl/min for *in vitro* samples (unpublished results). However, smaller columns have much less loading capacity and the system is considerably less robust. Therefore, we chose to run our analyses at more conventional flow rates, since sensitivity was sufficient to monitor all characterized HNE modifications from cell culture samples in a targeted manner.

Furthermore, the present study integrated a 2D-LC-MS/MS approach combining ion-exchange fractionation and RP chromatography. However, as

discussed in Section 1.2.3, HILIC is also commonly used for detection of modified peptides (Di Palma *et al.*, 2012). Due to the difference in mechanisms of separation, HILIC demonstrates remarkable orthogonality with RP chromatography and thus could be coupled with RP to enable MD separations. Previous works have demonstrated that HILIC-RP can increase the number of identified proteins compared to SCX-RP (Li *et al.*, 2013; Longworth *et al.*, 2012). However, despite its remarkable separation power and orthogonality of 2D HILIC combinations for shotgun proteomics analyses, HILIC-based platforms suffer from some technical challenges including solvent incompatibility between HILIC and other dimensions, and poor solubility of peptides in the highly organic solvents used in HILIC (Zhao *et al.*, 2012). Nonetheless, a HILIC-based MD-LC approach could be tested in the future to improve the detection efficiency of APAP-modified peptides from *in vivo* samples.

As discussed earlier (Section 7.1), the use of data-dependent acquisition can be another limitation to the method developed in this thesis. As demonstrated in Figure 7.7, horizontal selection of the precursor ions (usually by signal intensity) used in DDA (Figure 7.7a) can exclude less sensitive ions from MS/MS analysis and thus some potentially interesting peptides would not be sequenced. On the contrary, data-independent acquisition (DIA) operates via unbiased acquisition of consecutive survey scans and product ion spectra for all the precursors contained in predetermined vertical selection frames known as "isolation" or "selection windows" (Figure 7.7b). DIA thus can potentially provide higher sensitivity and greater reproducibility than DDA (Bern *et al.*, 2010). A more recent variation of DIA, SWATH[15], combines a sequential windowed acquisition with a post-acquisition targeted data analysis allowing for better quantitation of complex proteomics samples with higher accuracy and reproducibility (Gillet *et al.*, 2012). Therefore SWATH-MS analysis of the RLM samples can in theory lead to a more efficient detection of less-abundant and/or less-

[15] Sequential Window-Acquisition of All Theoretical Mass Spectra

sensitive APAP-modified peptides. However, DIA considerably increases the MS cycle time, and thus often requires longer LC gradients reducing the analysis throughput.

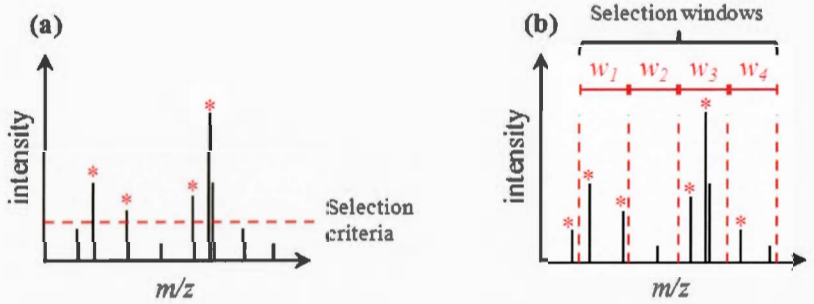


Figure 7.7 Graphical comparison between horizontal and vertical selection of precursor ions in DDA (**a**) and DIA (**b**). Selected ions for the dependent MS/MS are marked with an asterisk.

Finally, as described in Chapter 1, covalent adduct formation is not limited to proteins, and other biomolecules can be targeted by reactive metabolites, depending on the chemical and structural nature of the species involved. For example, RNA covalent modifications, such as methylation, inosination and polyadenylation, have been widely studied in recent years as they severely impair systematic characterization of the "epitranscriptome": the entire set of coding and noncoding RNA chemical modifications. These modifications are believed to influence parameters such as RNA stability, translation, trafficking, localization, enzymatic or sensing activity, regulatory capabilities, or patterns of interaction with other molecules. Current methods for identification of RNA modifications are mostly based on immunoprecipitation followed by RNA-sequencing, which offer low limit of detection and limited accuracy. LC-MS/MS methods have been recently introduced to this area enabling more sensitive, accurate, and robust measurements (Satterlee *et al.*, 2014). LC-MS/MS is also the method of choice for characterization and quantitation

of DNA covalent modifications in biological samples (Balbo *et al.*, 2014). DNA adduct detection strategies often include SPE or immunoaffinity enrichment, HILIC or CE separations, combined with MALDI or ESI-MS. Alkylation, oxidation, epoxidation, nitrosylation and cyclization are the most common DNA covalent modifications cause by reactive metabolites with potential mutagenic consequences (Tretyakova *et al.*, 2013). With appropriate sample preparation, the present approach can be adapted to identify covalent adducts of other macromolecules including phospholipids, RNA and DNA. Table 7.1 summarizes improvements that can be made to increase the efficiency and/or flexibility of the analytical method developed in this study.

Table 7.1 List of suggested modifications and extensions to this work for future studies

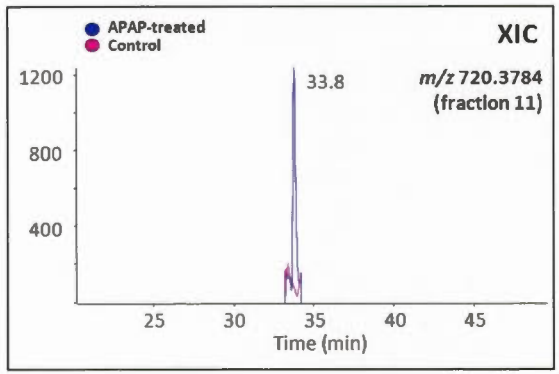
Conceptual Modifications and Improvements
Identification of protein adducts formed with other reactive metabolites (e.g. HNE)
Validation of the findings through *in vivo* studies
Application of affinity separation techniques for sample enrichment prior to the LC-MS/MS analysis
Method adaptation for the analysis of covalent modification of other macro-molecules (e.g. RNA, DNA)
Technical Enhancements
Incorporation of other orthogonal LC combinations to increase separation efficiency (e.g. HILIC-SCX, HILIC-RP)
Replacing DDA with DIA (or SWATH) to increase sensitivity for more targeted applications

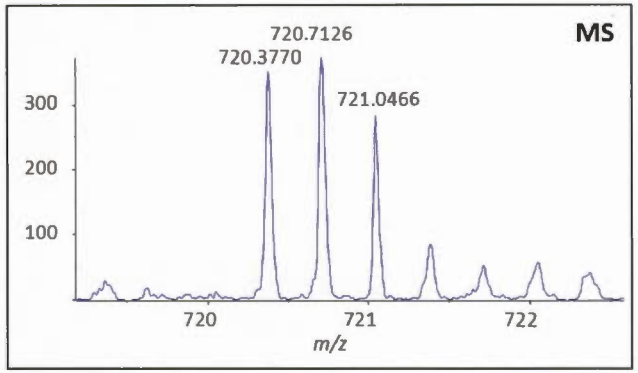
* * *

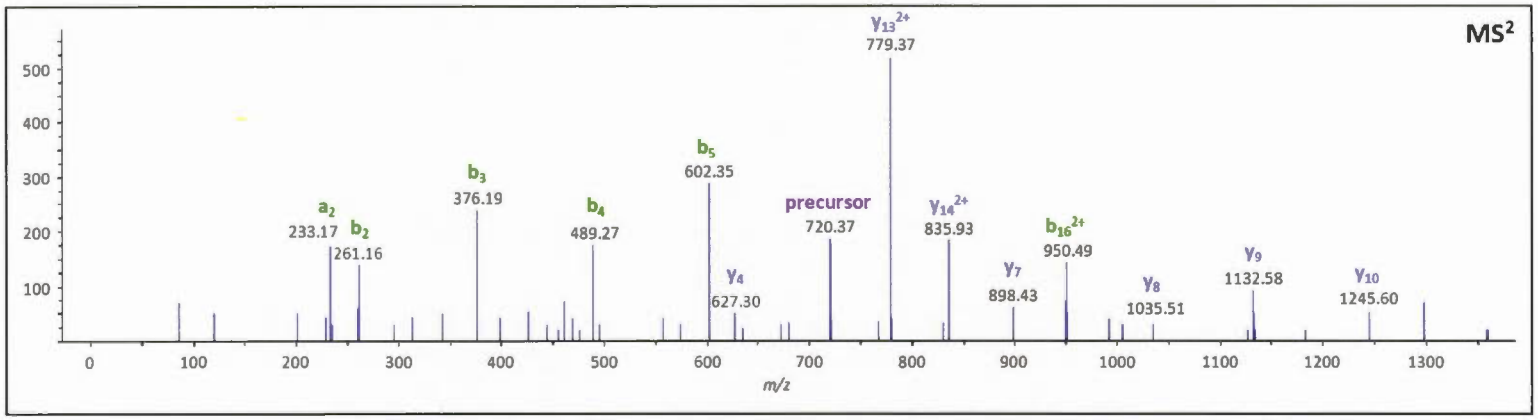
APPENDIX A

HIGH-RESOLUTION UHPLC-RP-MS/MS DATA FROM THE APAP-MODIFIED PEPTIDES IDENTIFIED IN RAT LIVER MICROSOMAL INCUBATIONS

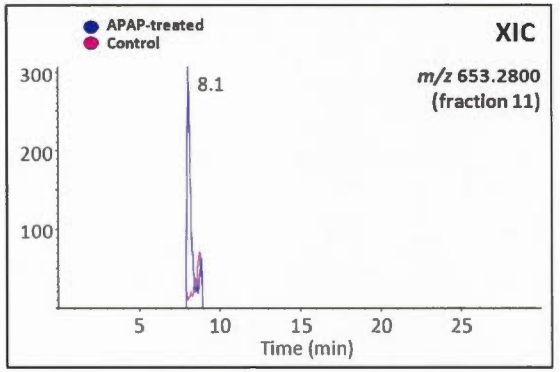
Peptide	Protein	*m/z*	RT (min)	Charge (*z*)
FIDLIPTNLPHAVTCDIK	CP2C6_RAT	720.3784	33.8	3

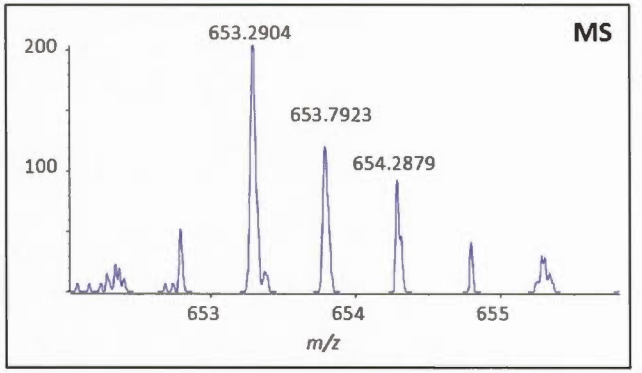


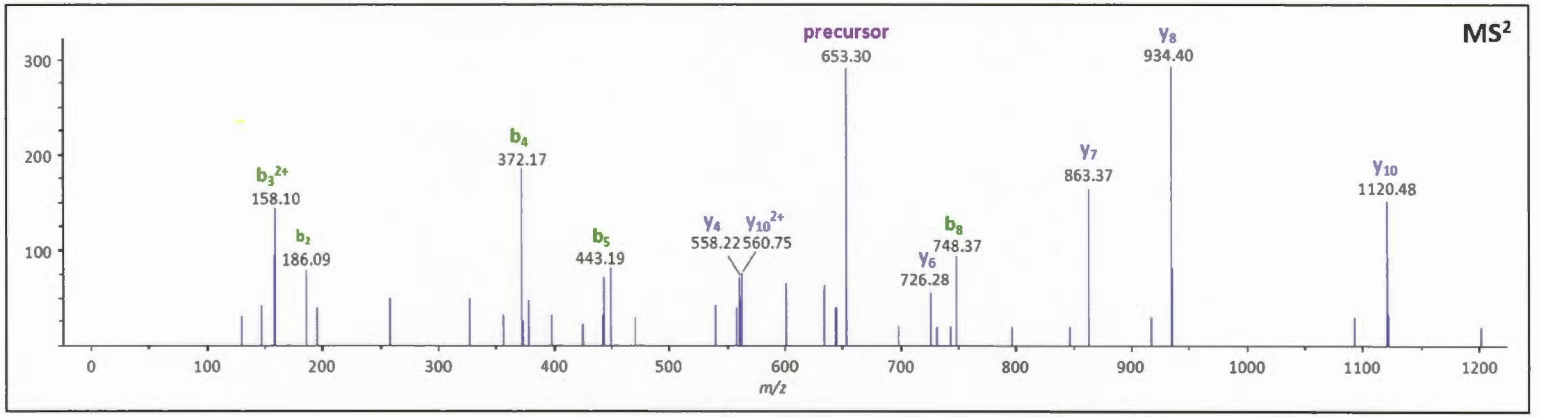




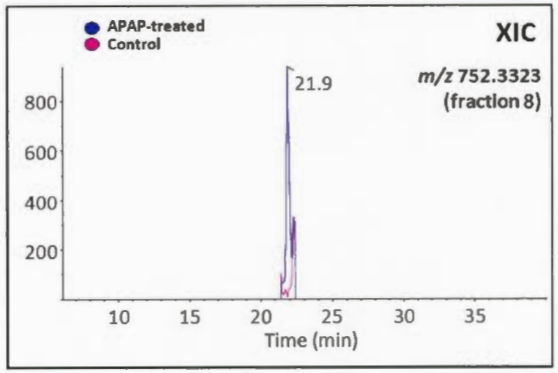
Peptide	Protein	*m/z*	RT (min)	Charge (*z*)
GKEGAHAPCASE	MADD_RAT	653.2800	8.1	2

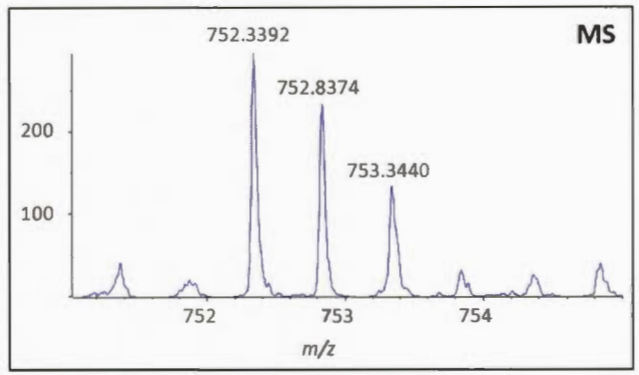


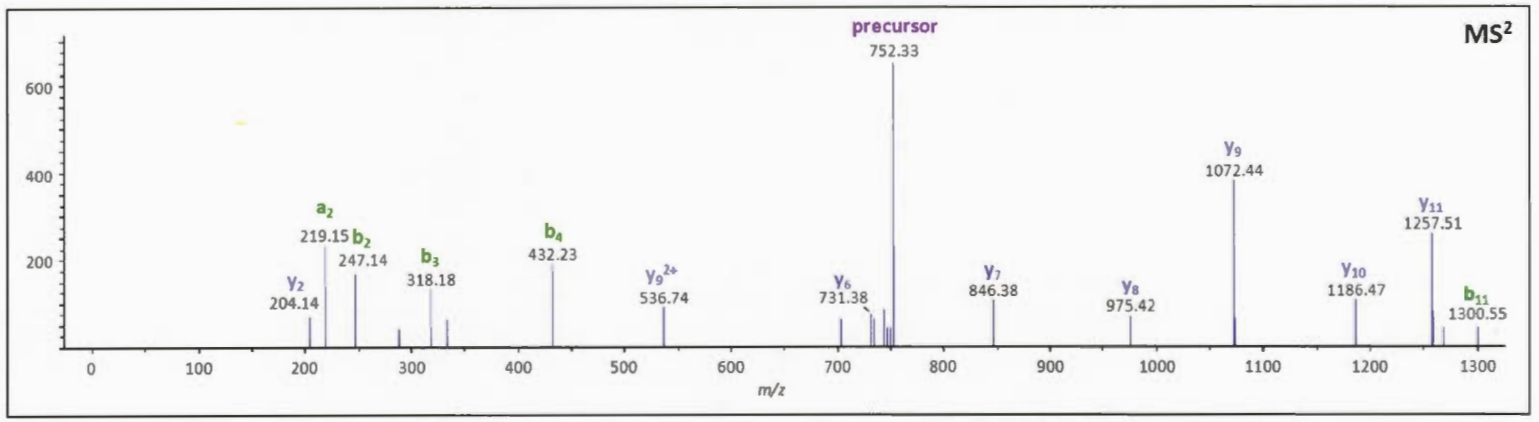




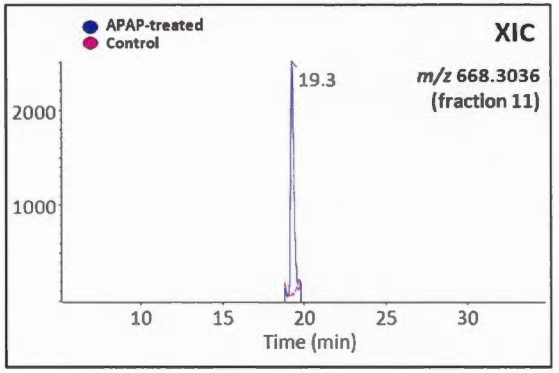
Peptide	Protein	*m/z*	RT (min)	Charge (z)
VFANPEDCAGFGK	MGST1_RAT	752.3323	21.9	2

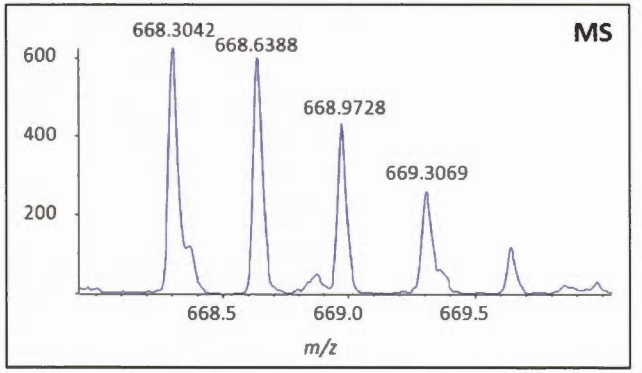


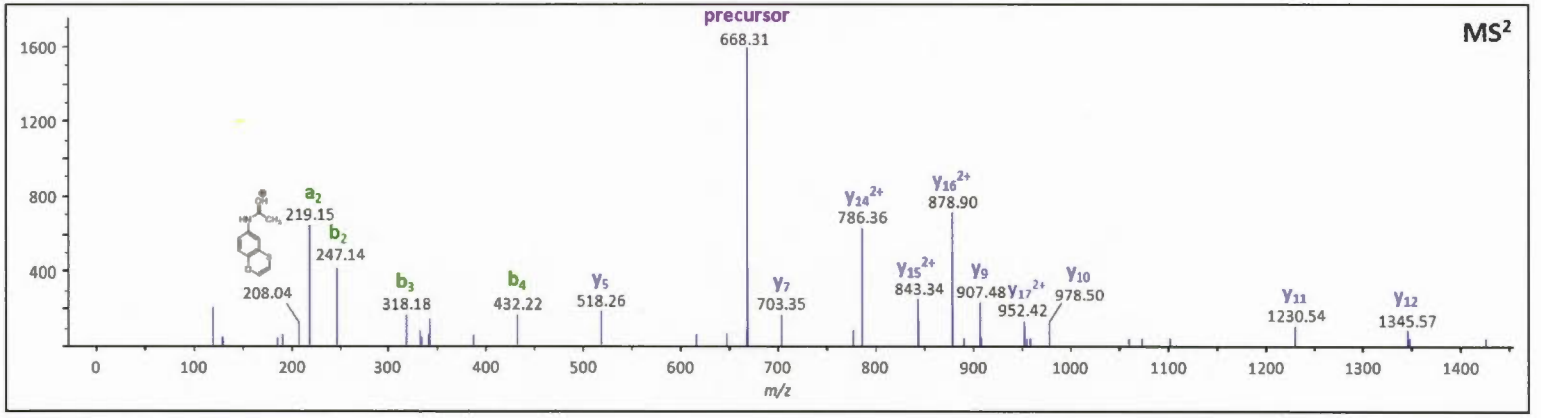




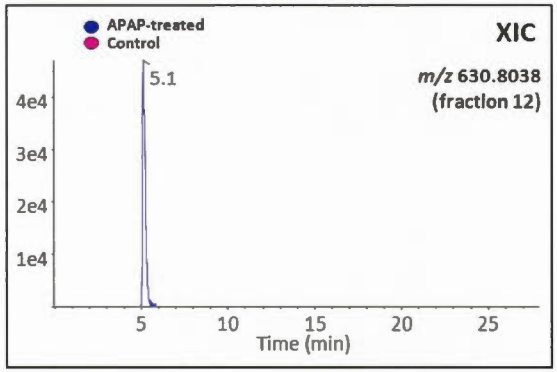
Peptide	Protein	*m/z*	RT (min)	Charge (z)
VFANPEDCAGFGKGENAK	MGST1_RAT	668.3036	19.3	3

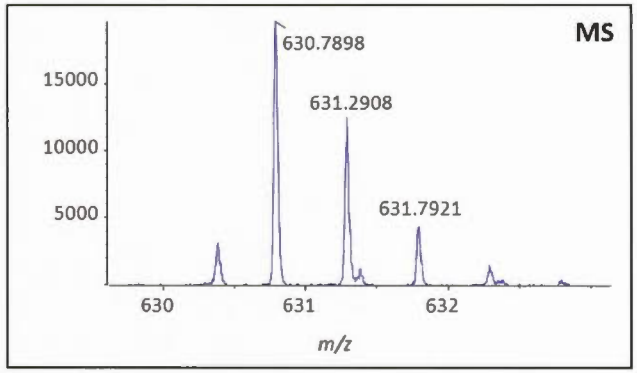


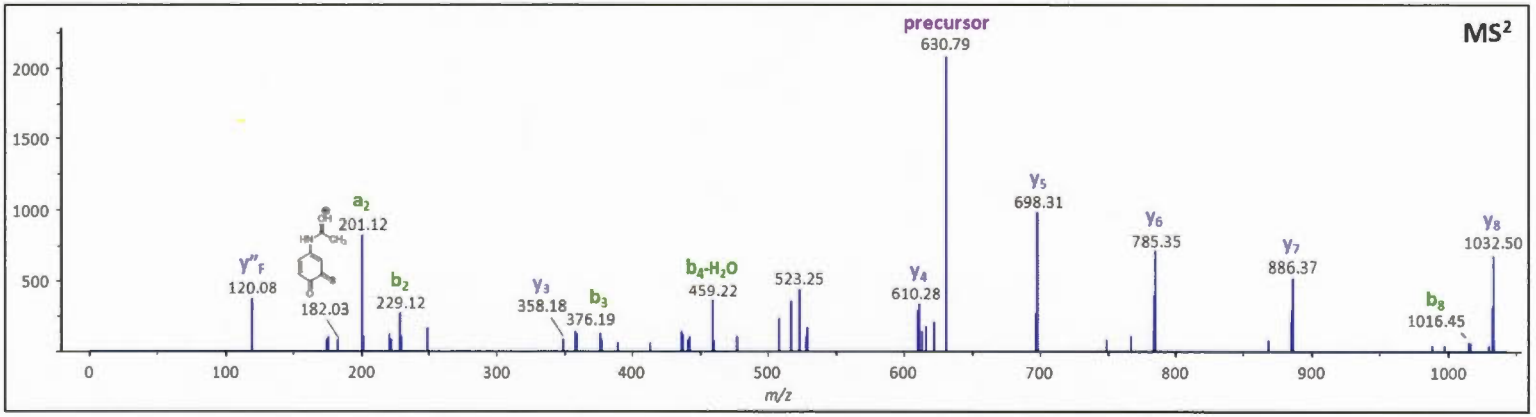




Peptide	Protein	*m/z*	RT (min)	Charge (z)
MPFTSSCLII	NU5M_RAT	630.8038	5.1	2



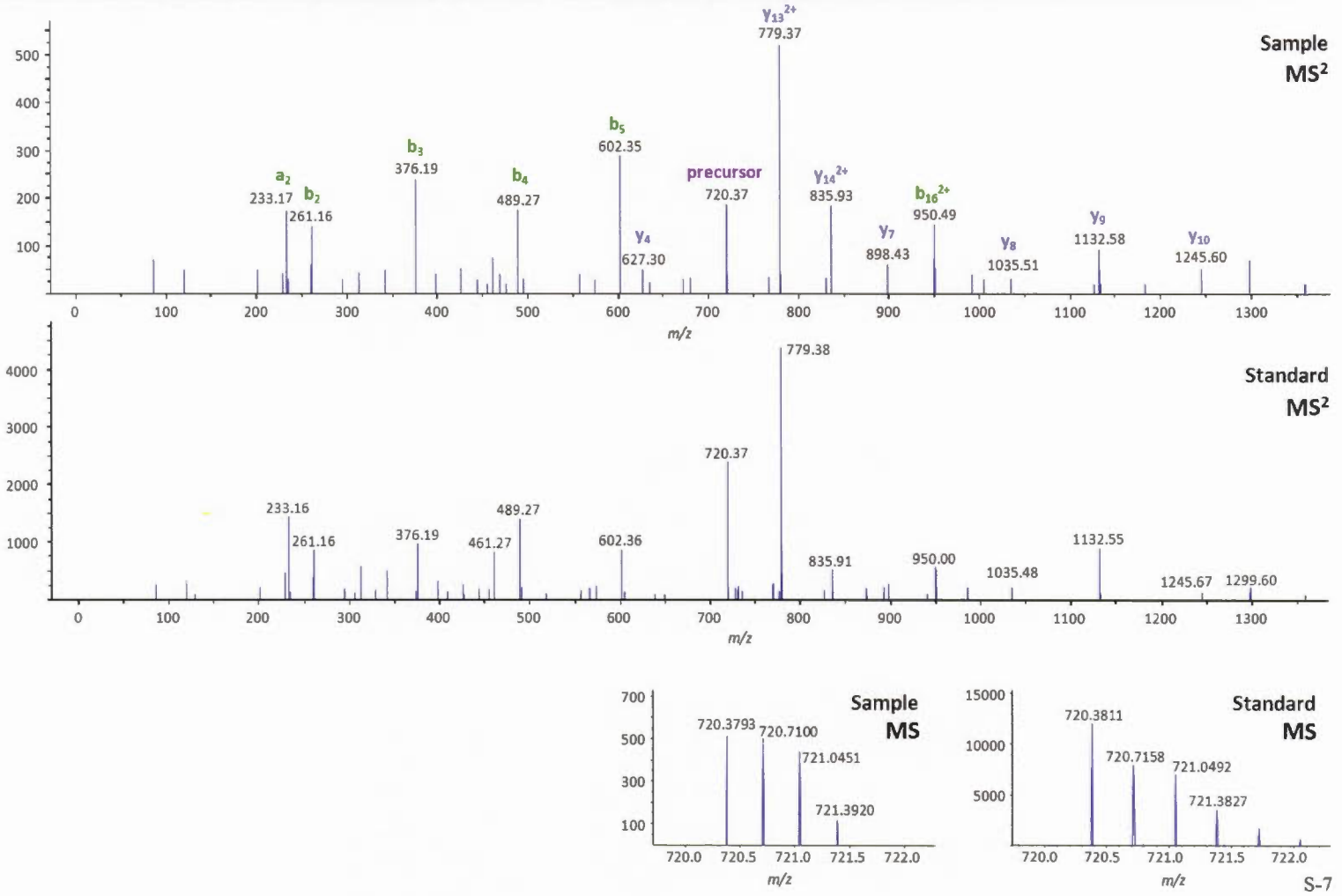




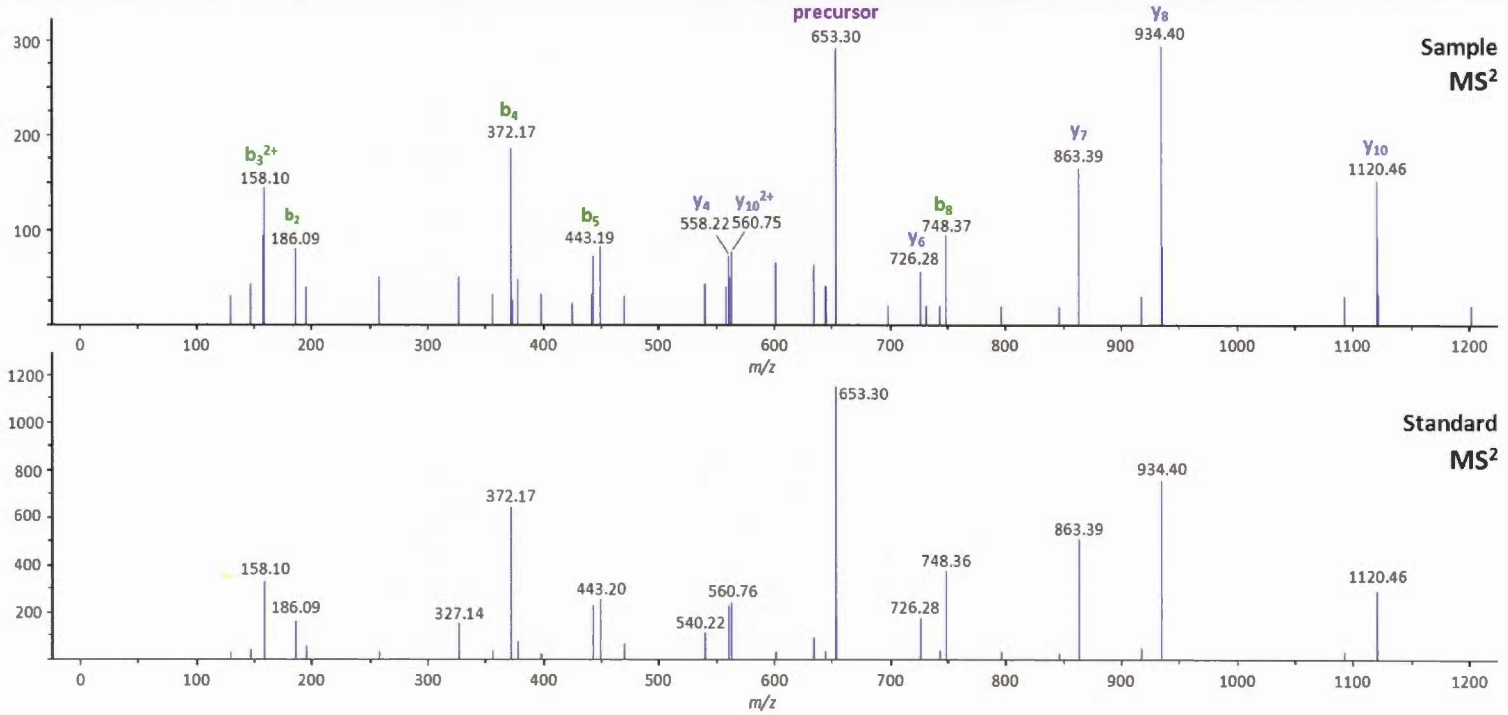
APPENDIX B

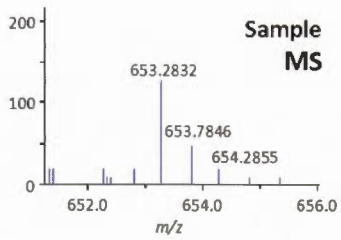
MS AND MS/MS SPECTRA FROM APAP-MODIFIED PEPTIDES IDENTIFIED IN RAT LIVER MICROSOMAL SAMPLES AND APAP-MODIFIED SYNTHETIC STANDARDS (TRAPPING EXPERIMENTS)

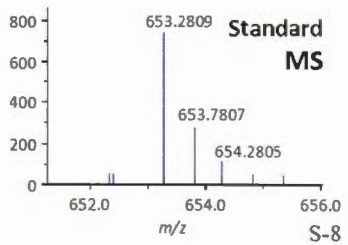
Peptide	Protein	*m/z*	RT (min)	Charge (*z*)
FIDLIPTNLPHAVTCDIK	CP2C6_RAT	720.3784	33.8	3



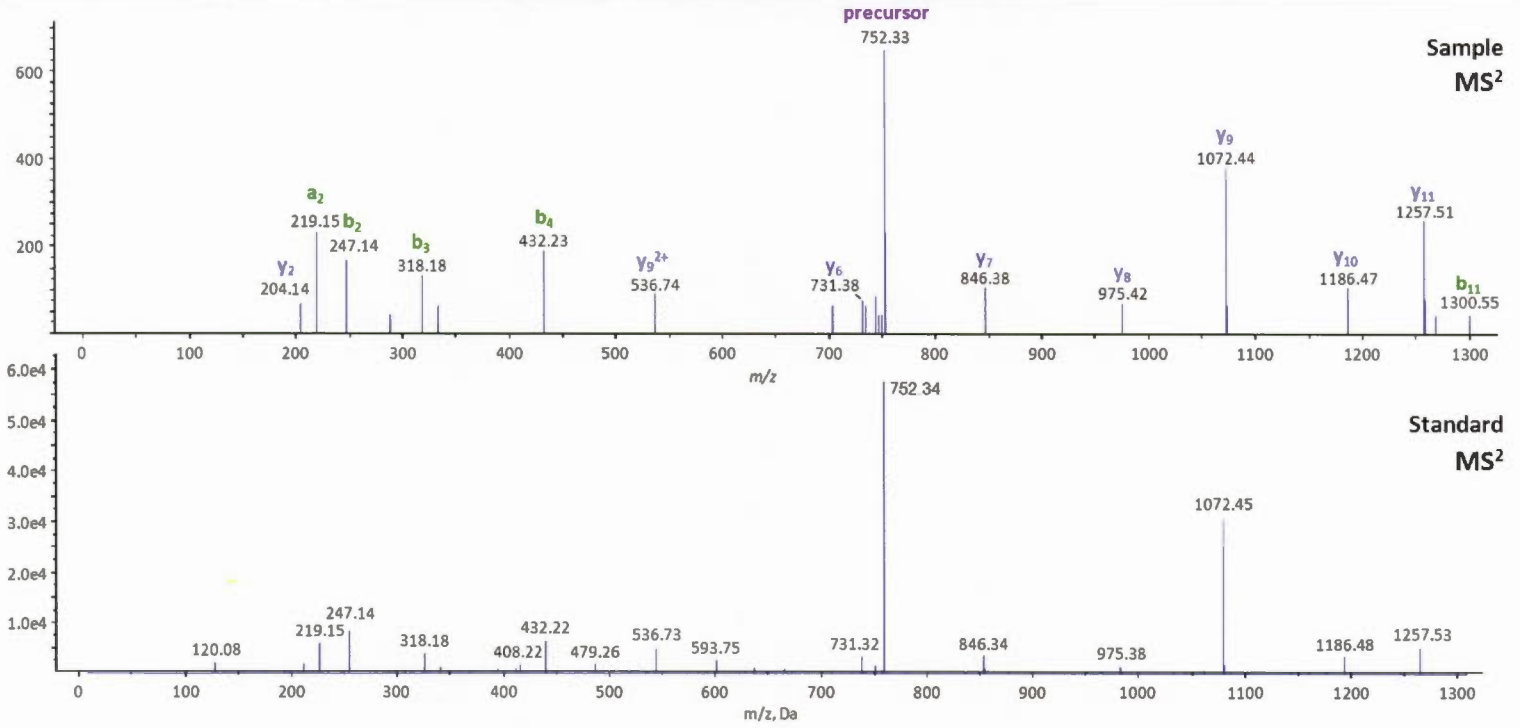
Peptide	Protein	*m/z*	RT (min)	Charge (*z*)
GKEGAHAPCASE	MADD_RAT	653.2800	8.1	2

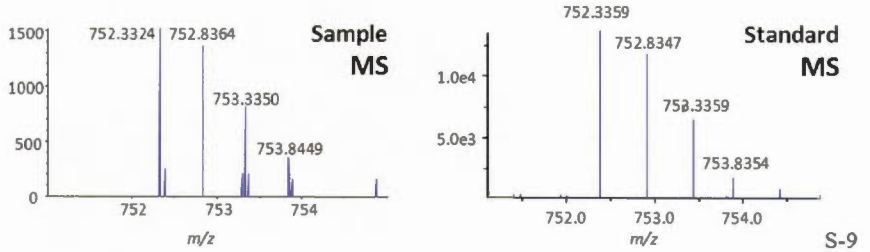




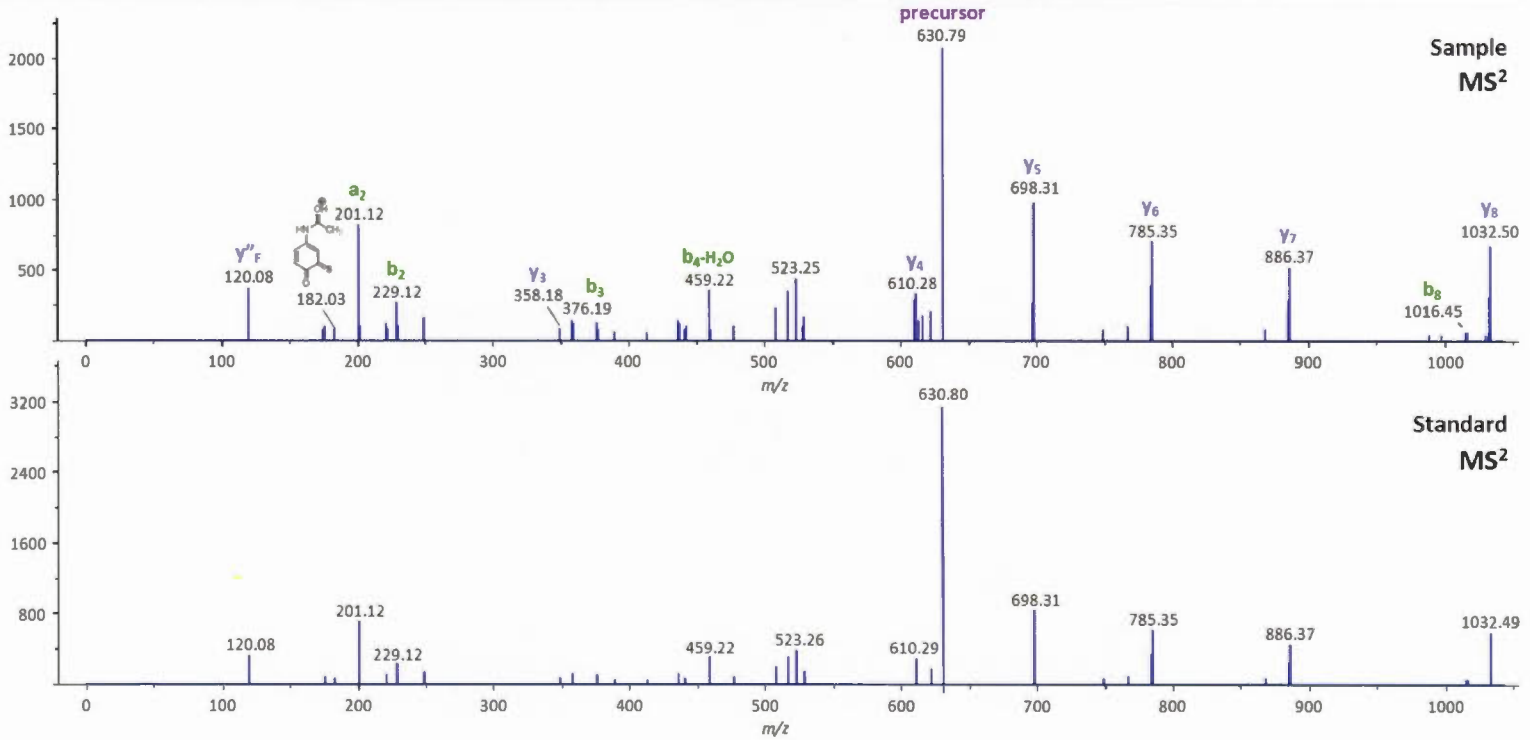


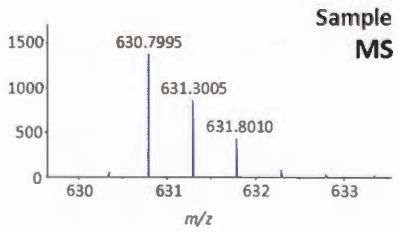
Peptide	Protein	*m/z*	RT (min)	Charge (*z*)
VFANPEDCAGFGK	MGST1_RAT	752.3323	21.9	2

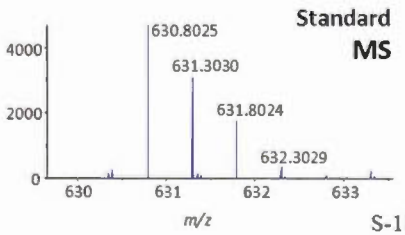




Peptide	Protein	*m/z*	RT (min)	Charge (z)
MPFTSSCLII	NU5M_RAT	630.8038	5.1	2



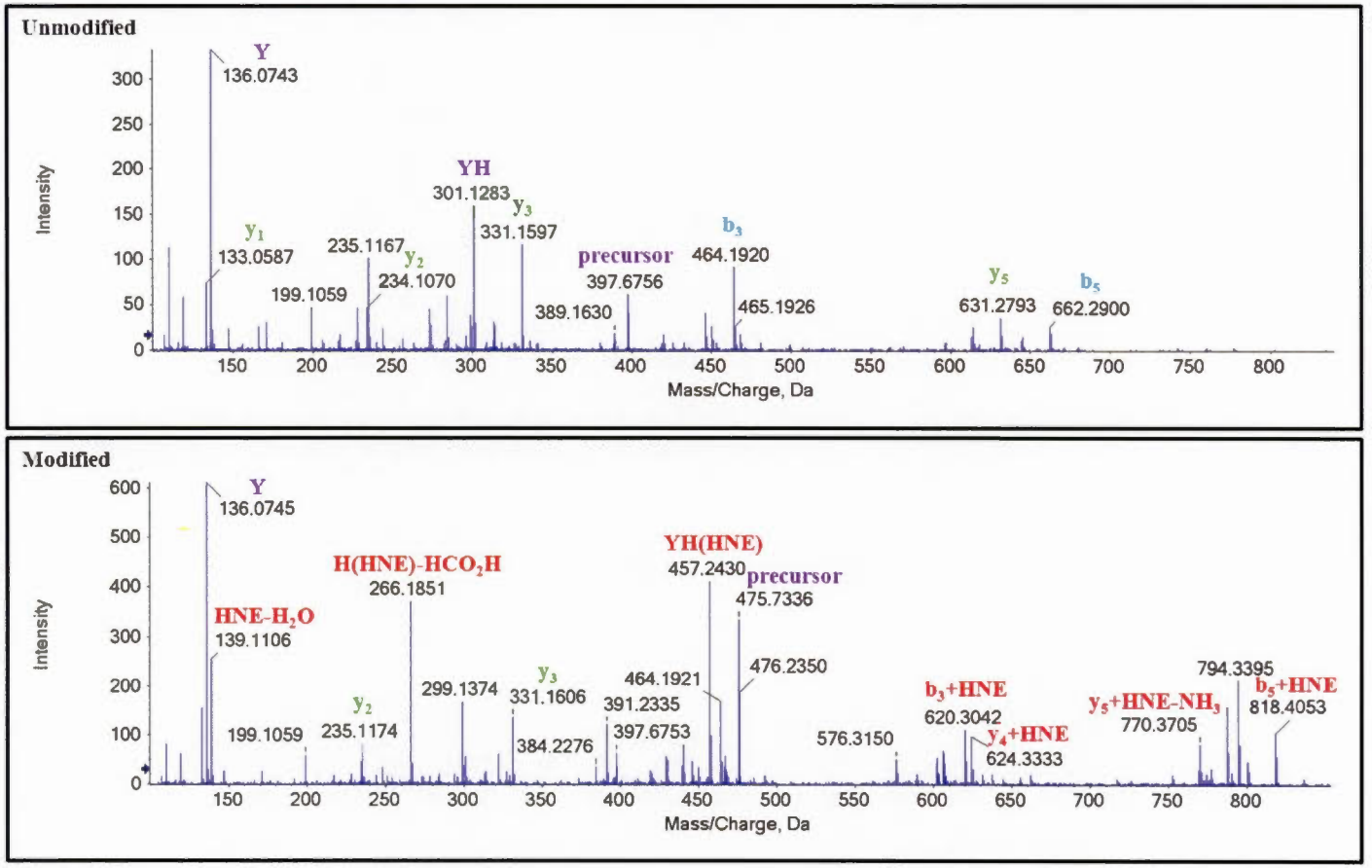




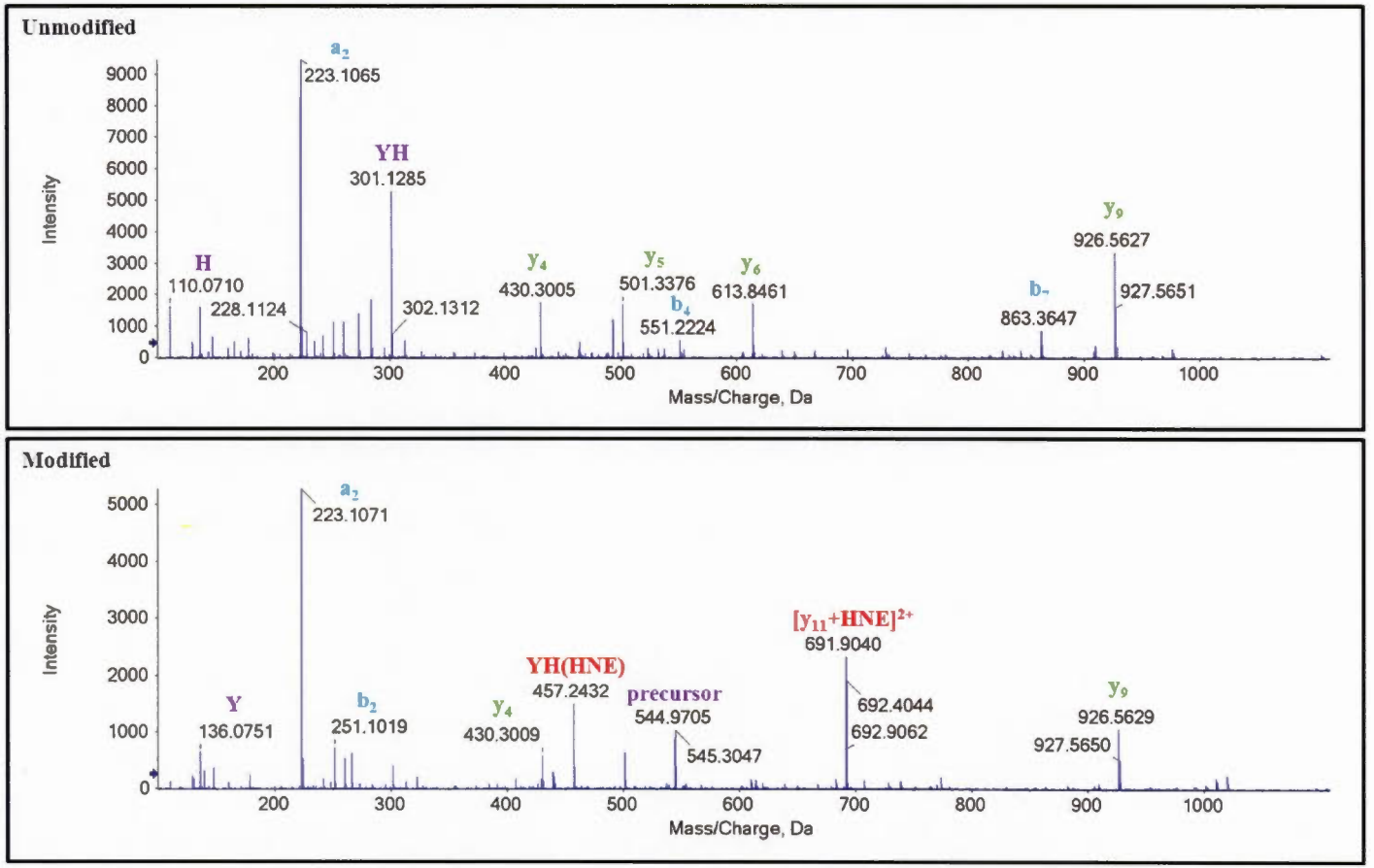
APPENDIX C

HIGH-RESOLUTION MS/MS SPECTRA OF THE UNMODIFIED AND HNE-MODIFIED PEPTIDES ASSOCIATED WITH EACH MMP-13 MODIFICATION SITE CHARACTERIZED IN *IN VITRO* SAMPLE

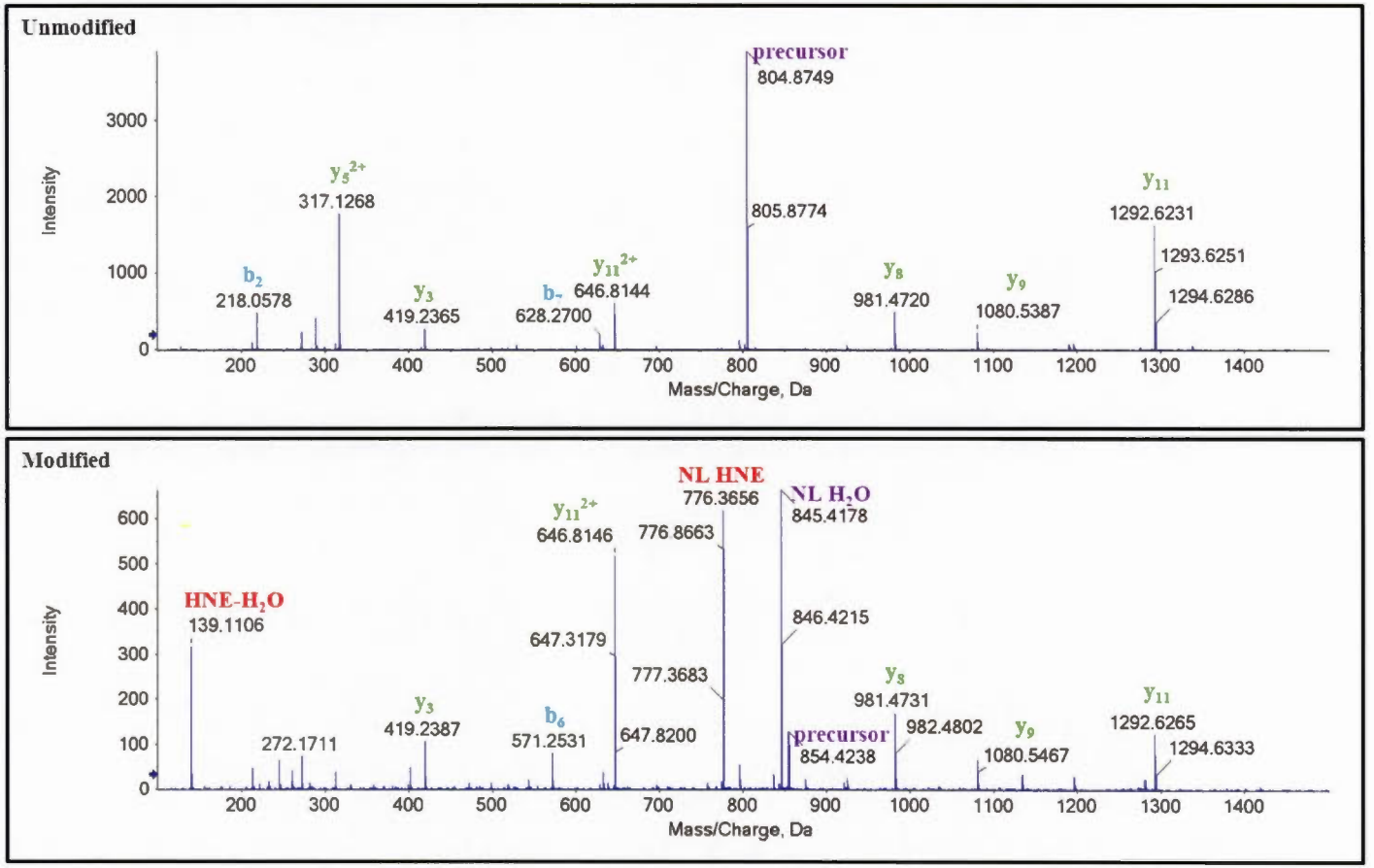
YYH[48]PTN	z	m/z	RT (min)
Unmodified	2	397.6771	2.7
HNE-modified	2	475.7346	9.1



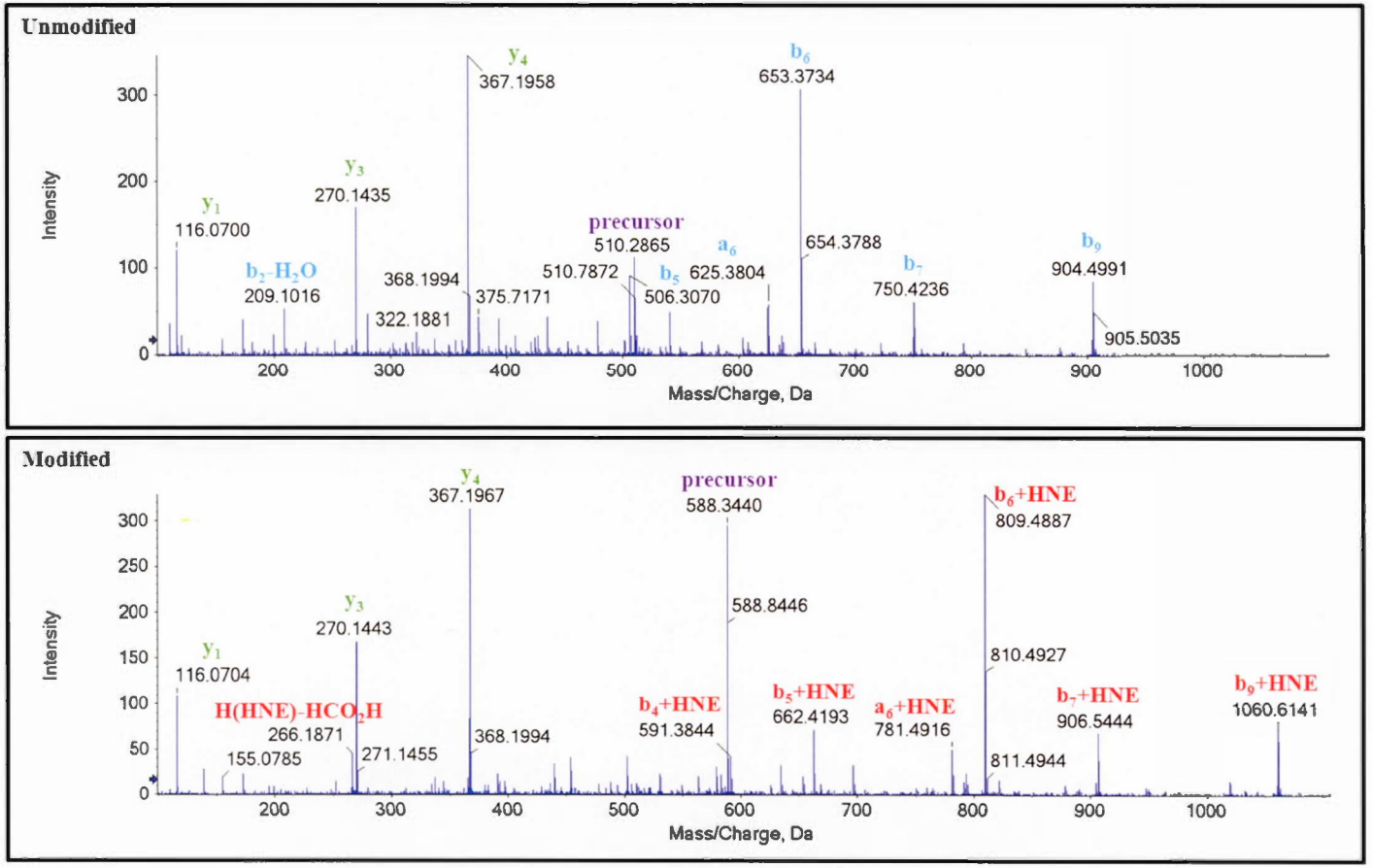
SYYH[48]PTNLAGILK	*z*	*m/z*	RT (min)
Unmodified	3	492.9330	11.1
HNE-modified	3	544.9714	13.3



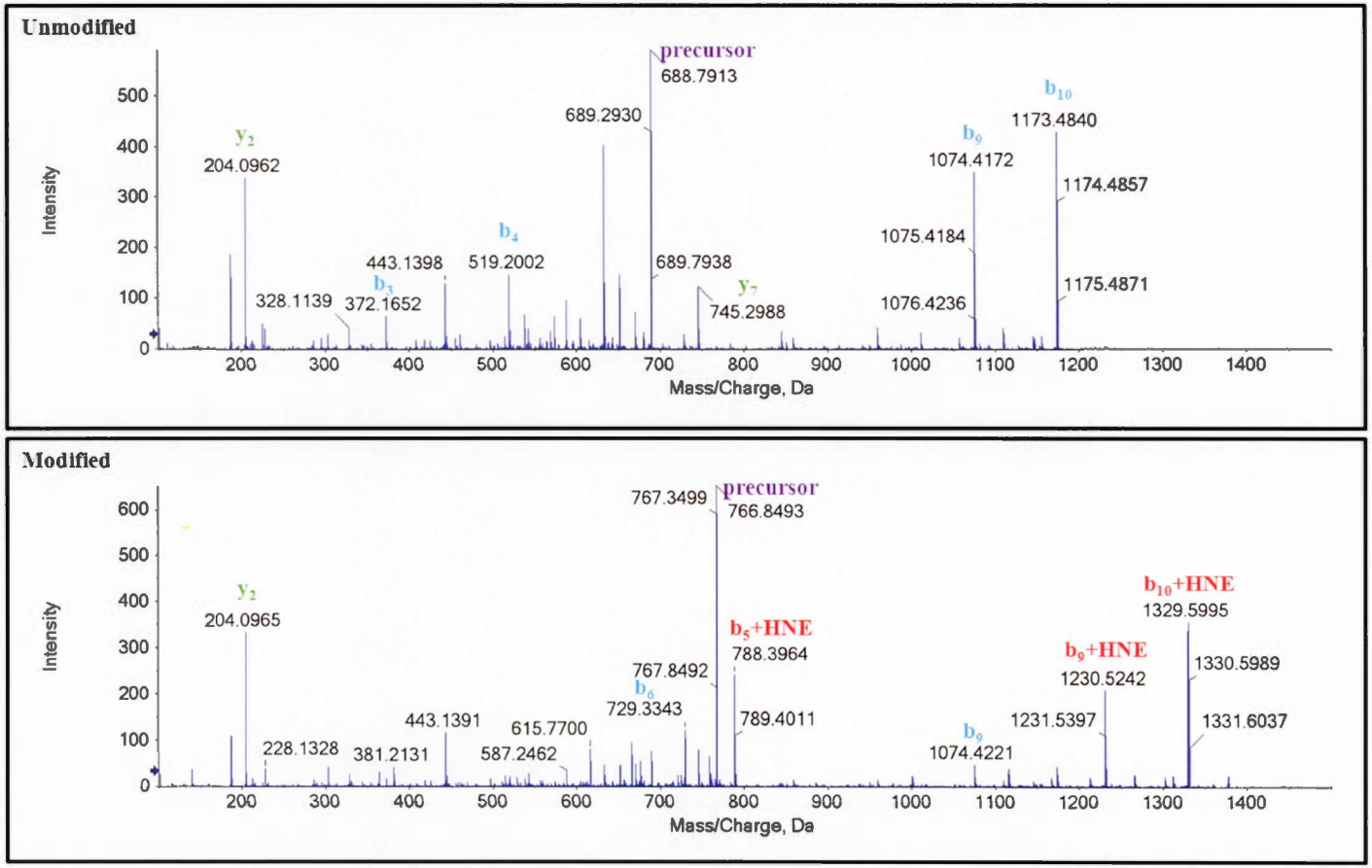
C^{96}GVPDVGEYNVFPR	*z*	*m/z*	RT (min)
Unmodified	2	804.8774	12.1
HNE-modified	2	854.4242	14.8



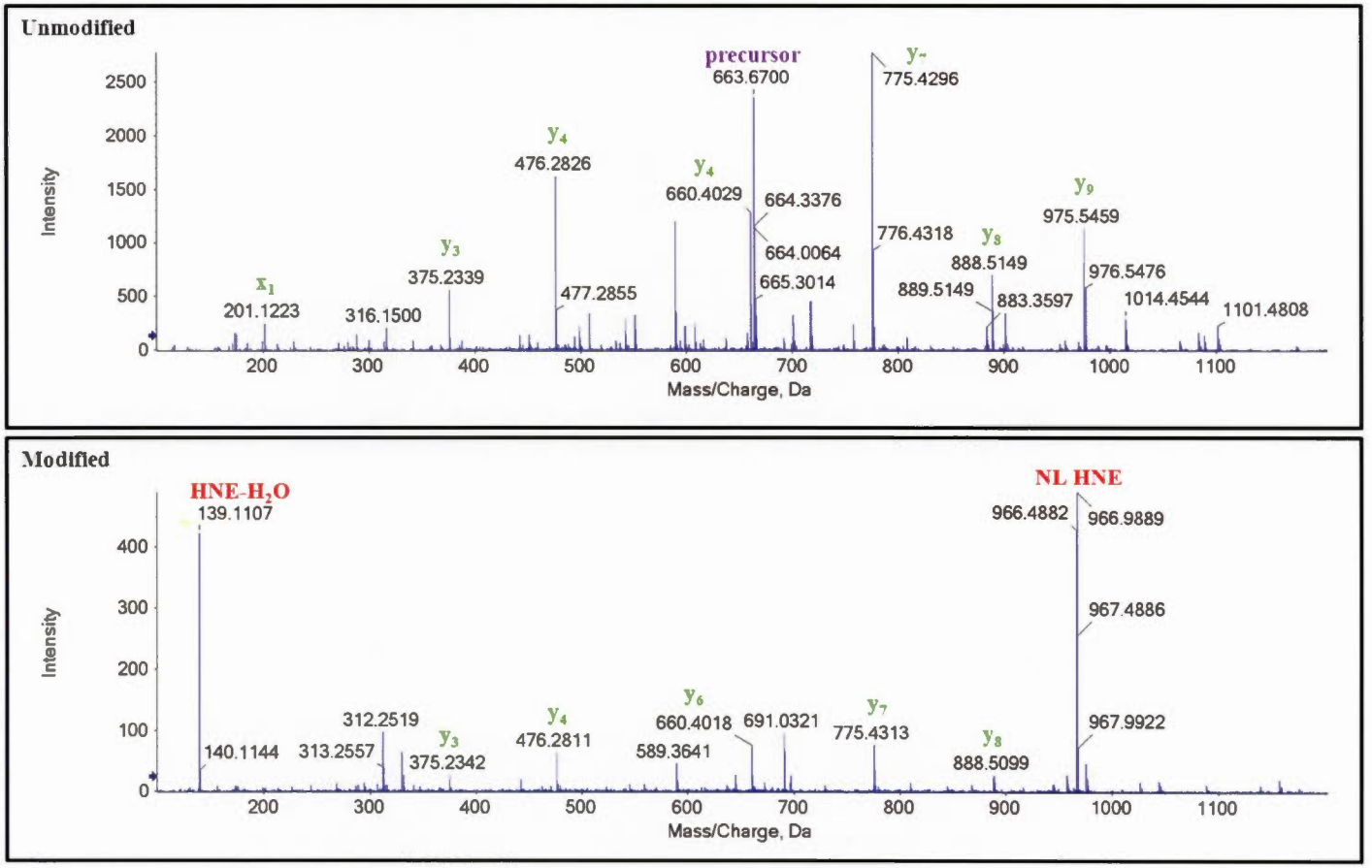
LLAH[187]AFPPGP	z	m/z	RT (min)
Unmodified	2	510.2873	10.4
HNE-modified	2	588.3448	13.7



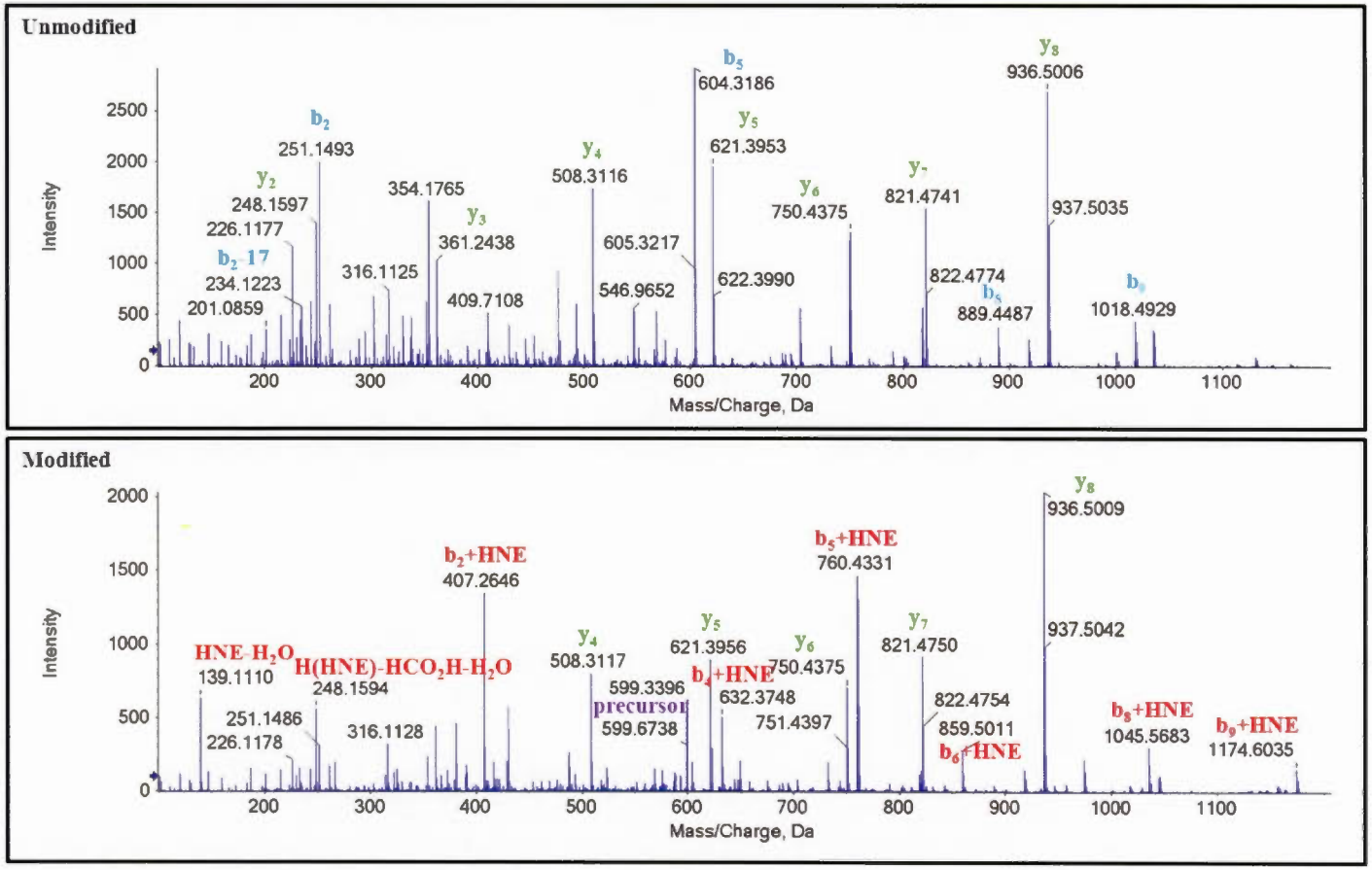
SH[251]FM(Ox)LPDDDVQG	z	m/z	RT (min)
Unmodified	2	688.7930	10.0
HNE-modified	2	766.8505	12.8



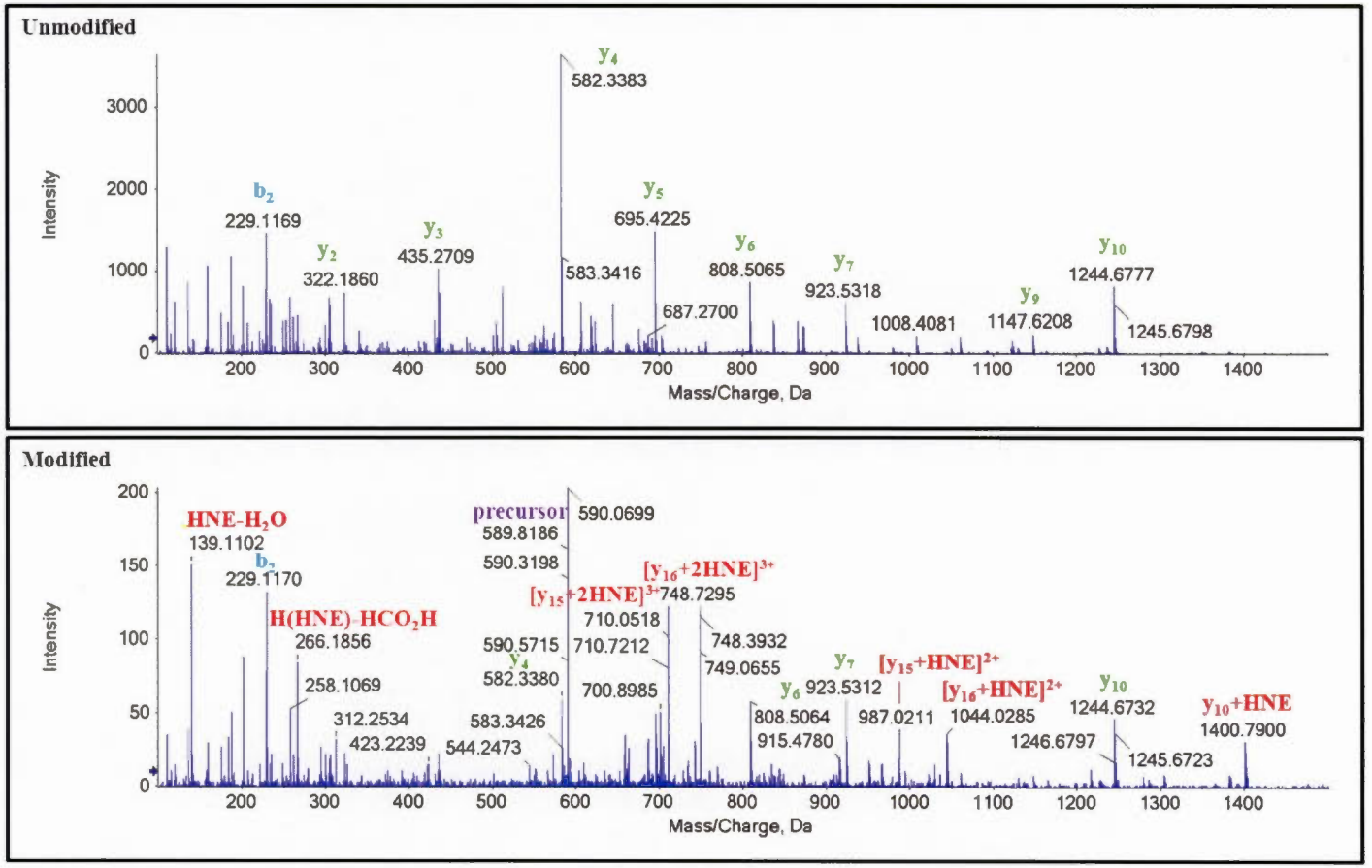
TPDKC[284]DPSLSLDAITSLR	z	m/z	RT (min)
Unmodified	3	663.6701	12.4
HNE-modified	3	696.7013	14.9



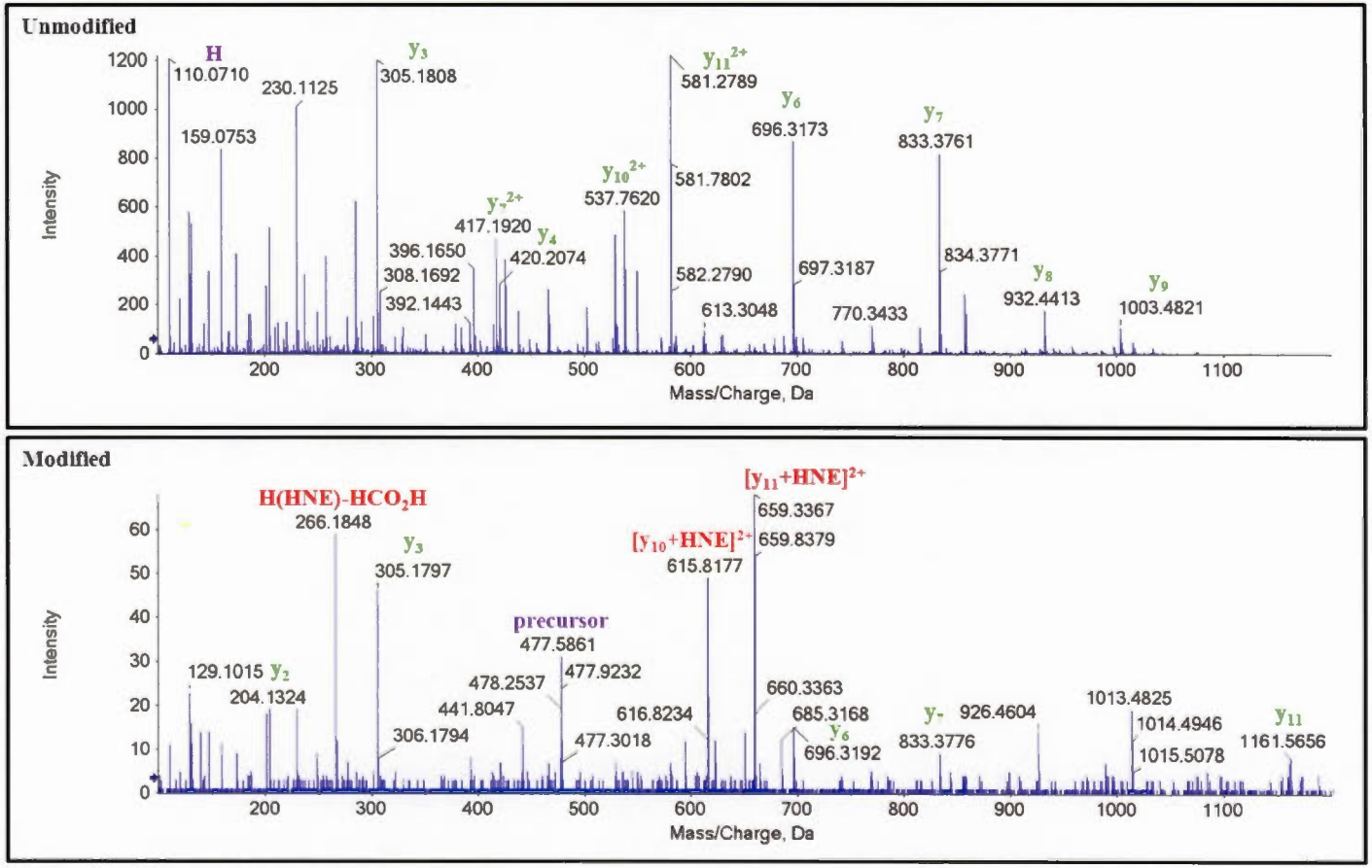
LH[312]PQQVDAELFLTK	*z*	*m/z*	RT (min)
Unmodified	3	546.9665	11.4
HNE-modified	3	599.0049	13.2



IDAAYEH[340]PSH[343]DLIFIFR	*z*	*m/z*	RT (min)
Unmodified	4	511.7626	13.0
HNE-modified	4	589.8202	16.9



ISAAVH[387]FEDTGK	z	m/z	RT (min)
Unmodified	3	425.5507	8.1
HNE-modified	3	477.5891	10.8



APPENDIX D

MASSBOX PEAK PICKER SOURCE CODE

"Peak-picking" is often employed in LC-MS-based "omics" for data reduction and filtering using LC-MS signal-related parameters. Several commercial and free software packages are available for this purpose based on different search, or picking, algorithms (Rafiei and Sleno, 2015). In Chapter 5, an in-house custom-built open-source application (MassBox) was employed for peak-picking based on the exact mass difference and retention time shift between LC-MS peak pairs. MassBox was originally developed during this dissertation for routine MS-based omics applications using Java programming language (JDK version 1.7) and Eclipse Kepler (Eclipse Foundation, Ottawa, ON, Canada) integrated development environment (IDE). MassBox is a package of six operational modules including molecular weight (MW) calculator, neutral/ion converter, protein MW and GRAVY calculator, formula finder, LC-MS data redundancy remover, and peak-picker. The package is freely available for download via SourceForge public source code repository at http://massbox.sourceforge.net.

MassBox Peak Picker employs a generic double comparison algorithm where *m/z* and retention time (RT) values are compared between two peak lists, normally from a control and treated sample. Peak pairs with a given mass difference within a certain mass error threshold (in ppm) and RT window will be reported. To remove potential false positives within each data set, *m/z* and RT values are only compared between the two lists, and each list is not compared to itself. The source code for this

module is included below. By convention, Java code comments are placed after // tags, or enclosed within /** */ delimiters.

```
import java.awt.Cursor;
import java.awt.event.ActionEvent;
import java.awt.event.ActionListener;
import java.beans.PropertyChangeEvent;
import java.beans.PropertyChangeListener;
import java.text.DecimalFormat;
import javax.swing.JFrame;
import javax.swing.JPanel;
import javax.swing.border.EmptyBorder;
import javax.swing.JLabel;
import javax.swing.JOptionPane;
import javax.swing.JScrollPane;
import javax.swing.JTextArea;
import javax.swing.JTextField;
import javax.swing.JButton;
import javax.swing.JProgressBar;
import javax.swing.ScrollPaneConstants;
import javax.swing.SwingUtilities;
import javax.swing.SwingWorker;
import java.awt.Toolkit;
import javax.swing.border.BevelBorder;


public class frmPeakPicker extends JFrame implements ActionListener,
PropertyChangeListener {

    /** PeakPicker class definition */

    private JPanel pnlMain;
    private JTextField txtDeltamass, txtErrorThreshold, txtRTWidth;
    private JTextArea txtList1, txtList2, txtRT1, txtRT2,
                      txtResults;
    private JButton btnSearch;
    private JProgressBar progressBar;
    private JLabel lblSearching;
    private Task task;
    private int  progress;
    private long hits;
    private long startTime, elapsedTime;

    /** Task to be done in background */

    class Task extends SwingWorker<Void, Void> {

        @Override
        public Void doInBackground() {
```

```
// To verify the input

if (!isDouble(txtDeltamass.getText()) ||
    !isDouble(txtErrorThreshold.getText()) ||
    !isDouble(txtRTWidth.getText())) {
      JOptionPane.showMessageDialog(null,
            "Unexpected search parameters!");
      return null;
}

// To initialize the search

double delta  =
  Math.abs(Double.valueOf(txtDeltamass.getText()));
double er_thr =
  Math.abs(Double.valueOf(txtErrorThreshold.getText()));
double rt_thr =
  Math.abs(Double.valueOf(txtRTWidth.getText()));
double upperLimit = delta + er_thr / 1e6;
double lowerLimit = delta - er_thr / 1e6;
double mz_1, mz_2, rt_1, rt_2, dmz, drt;
boolean match_mz, match_rt;
DecimalFormat mzFormat = new DecimalFormat("0.0000");
DecimalFormat rtFormat = new DecimalFormat("0.0");

hits = 0;

String mz1List = txtList1.getText();
String rt1List = txtRT1.getText();
String mz2List = txtList2.getText();
String rt2List = txtRT2.getText();

int mz1Lines, mz2Lines, rt1Lines, rt2Lines;
mz1Lines = mz2Lines = rt1Lines = rt2Lines = 0;

for (String line : mz1List.split("\\n")) mz1Lines++;
for (String line : rt1List.split("\\n")) rt1Lines++;
for (String line : mz2List.split("\\n")) mz2Lines++;
for (String line : rt2List.split("\\n")) rt2Lines++;

if (mz1Lines != rt1Lines || mz2Lines != rt2Lines) {
    JOptionPane.showMessageDialog(null,
"Number of peaks and retention times do not match!");
    return null;
}

if (mz1List.charAt(mz1List.length()-1) > 0) mz1List +=
      "\n";
if (rt1List.charAt(rt1List.length()-1) > 0) rt1List +=
      "\n";
if (mz2List.charAt(mz2List.length()-1) > 0) mz2List +=
      "\n";
```

```
if (rt2List.charAt(rt2List.length()-1) > 0) rt2List +=
      "\n";

String [] mz1 = new String [mz1Lines];
String [] rt1 = new String [rt1Lines];
String [] mz2 = new String [mz2Lines];
String [] rt2 = new String [rt2Lines];

String delim = "\n";

int i = 0, j = 0, k = 0;

while (j >= 0) { // To initiate mz1 list
    j = mz1List.indexOf(delim, i);
    if (j == i ) break;
    if (j >= 0) {
        mz1 [k] = mz1List.substring(i, j);
        i = j + delim.length();
    }
    k++;
}
i = j = k = 0;
while (j >= 0) { // To initiate rt1 list
    j = rt1List.indexOf(delim, i);
    if (j == i ) break;
    if (j >= 0) {
        rt1 [k] = rt1List.substring(i, j);
        i = j + delim.length();
    }
    k++;
}
i = j = k = 0;
while (j >= 0) { // To initiate mz2 list
    j = mz2List.indexOf(delim, i);
    if (j == i ) break;
    if (j >= 0) {
        mz2 [k] = mz2List.substring(i, j);
        i = j + delim.length();
    }
    k++;
}
i = j = k = 0;
while (j >= 0) { // To initiate rt2 list
    j = rt2List.indexOf(delim, i);
    if (j == i ) break;
    if (j >= 0) {
        rt2 [k] = rt2List.substring(i, j);
        i = j + delim.length();
    }
    k++;
}
```

```
        // To initialize task's progress property

        long offset = 1;
        long maxHits = mz1Lines * mz2Lines;
        progress = 0;
        setProgress(0);
        progressBar.setValue(0);
        startTime = System.currentTimeMillis();

        // To do in background (the peak-picking operation)

        while (progress < 100) {
            for (int m = 0; m < mz1Lines; m++) {
                for (int n = 0; n < mz2Lines; n++) {
                    if (isDouble(mz1 [m]) &&
                        isDouble(mz2 [n]) &&
                        isDouble(rt1 [m]) &&
                        isDouble(rt2 [n])) {
                          mz_1 = Double.valueOf(mz1 [m]);
                          mz_2 = Double.valueOf(mz2 [n]);
                          rt_1 = Double.valueOf(rt1 [m]);
                          rt_2 = Double.valueOf(rt2 [n]);
                          dmz  = Math.abs(mz_1 - mz_2);
                          drt  = Math.abs(rt_1 - rt_2);
                          match_mz = dmz < upperLimit &&
                                     dmz > lowerLimit;
                          match_rt = drt < rt_thr;
                          if (match_mz && match_rt) {
        txtResults.append(mzFormat.format(mz_1) + " (" +
                          rtFormat.format(rt_1) + " min) & " +
                          mzFormat.format(mz_2) + " (" +
                          rtFormat.format(rt_2) + " min) \n");
                            hits++;
                          }
                          offset++;
                          setElapsedTime();
                          progress = (int)(offset * 100 /
                                maxHits);
                          setProgress(progress);
                    }
                }
            }
        }
        return null;
    }

    /** Calculate elapsed time */

    private void setElapsedTime() {
        elapsedTime = (System.currentTimeMillis() - startTime) /
              1000;
        if (elapsedTime < 60) {
```

```
                progressBar.setString("Elapsed time : " +
                        Math.round(elapsedTime) + " sec");
                return;
            }
            else if (elapsedTime < 3600) {
                progressBar.setString("Elapsed time : " +
                        Math.round(elapsedTime/60) + " min");
                return;
            }
            else if (elapsedTime < 86400) {
                progressBar.setString("Elapsed time : " +
                        Math.round(elapsedTime/3600) + " h " +
                        Math.round((elapsedTime/3600-
                        Math.round(elapsedTime/3600))*60) + " min");
                return;
            }
            else {
                progressBar.setString("Elapsed time : " +
                        Math.round(elapsedTime/86400) + " d " +
                        Math.round((elapsedTime/86400-
                        Math.round(elapsedTime/86400))*24) + " h " +
                        Math.round((elapsedTime/3600-
                        Math.round(elapsedTime/3600))*60) + " min");
                return;
            }
        }

        /** While task is completed */

        @Override
        public void done() {
            Toolkit.getDefaultToolkit().beep();
            btnSearch.setEnabled(true);
            setCursor(null);
            lblSearching.setText(hits + " peak pair(s) found!");
        }
    }

    private boolean isDouble(String input) {
        try { Double.parseDouble(input); }
        catch(NumberFormatException e) { return false; }
        return true;
    }

    /** Launch the application */

    public static void main(String[] args) {
        SwingUtilities.invokeLater(new Runnable() {
            public void run() {
                try {
                    frmPeakPicker frame = new frmPeakPicker();
                    frame.setVisible(true);
```

```
                } catch (Exception e) {
                    e.printStackTrace();
                }
            }
        });
    }

    /** Create the frame */

    public frmPeakPicker() {
        setResizable(false);
        setTitle("Peak Picker");
        setDefaultCloseOperation(JFrame.DISPOSE_ON_CLOSE);
        setBounds(100, 100, 930, 404);
        pnlMain = new JPanel();
        pnlMain.setBorder(new EmptyBorder(5, 5, 5, 5));
        setContentPane(pnlMain);
        pnlMain.setLayout(null);

        JPanel pnlList1 = new JPanel();
        pnlList1.setBorder(new BevelBorder(BevelBorder.LOWERED,
            null, null, null, null));
        pnlList1.setBounds(10, 32, 292, 271);
        pnlMain.add(pnlList1);
        pnlList1.setLayout(null);

        txtList1 = new JTextArea();
        txtList1.setBounds(10, 32, 130, 226);
        JScrollPane scrollList1 = new JScrollPane(txtList1);
        scrollList1.setBounds(10, 32, 130, 226);
        scrollList1.setVerticalScrollBarPolicy(ScrollPaneConstants.
            VERTICAL_SCROLLBAR_ALWAYS);
        pnlList1.add(scrollList1);

        txtRT1 = new JTextArea();
        txtRT1.setBounds(150, 32, 130, 226);
        JScrollPane scrollRT1 = new JScrollPane(txtRT1);
        scrollRT1.setBounds(150, 32, 130, 226);
        scrollRT1.setVerticalScrollBarPolicy(ScrollPaneConstants.
            VERTICAL_SCROLLBAR_ALWAYS);
        pnlList1.add(scrollRT1);

        JLabel lblLis1 = new JLabel("Peak List 1 (control):");
        lblLis1.setBounds(10, 10, 147, 14);
        pnlMain.add(lblLis1);

        JLabel lblMZ1 = new JLabel("m/z:");
        lblMZ1.setBounds(10, 11, 147, 14);
        pnlList1.add(lblMZ1);

        JLabel lblRT1 = new JLabel("Retention time (min):");
        lblRT1.setBounds(150, 11, 147, 14);
```

```
pnlList1.add(lblRT1);

JPanel pnlList2 = new JPanel();
pnlList2.setLayout(null);
pnlList2.setBorder(new BevelBorder(BevelBorder.LOWERED,
    null, null, null, null));
pnlList2.setBounds(312, 32, 292, 271);
pnlMain.add(pnlList2);

txtList2 = new JTextArea();
txtList2.setBounds(10, 32, 130, 226);
JScrollPane scrollList2 = new JScrollPane(txtList2);
scrollList2.setBounds(10, 32, 130, 226);
scrollList2.setVerticalScrollBarPolicy(ScrollPaneConstants.
    VERTICAL_SCROLLBAR_ALWAYS);
pnlList2.add(scrollList2);

txtRT2 = new JTextArea();
txtRT2.setBounds(150, 32, 130, 226);
JScrollPane scrollRT2 = new JScrollPane(txtRT2);
scrollRT2.setBounds(150, 32, 130, 226);
scrollRT2.setVerticalScrollBarPolicy(ScrollPaneConstants.
    VERTICAL_SCROLLBAR_ALWAYS);
pnlList2.add(scrollRT2);

JLabel lblList2 = new JLabel("Peak List 2 (sample):");
lblList2.setBounds(312, 10, 147, 14);
pnlMain.add(lblList2);

JLabel lblMZ2 = new JLabel("m/z:");
lblMZ2.setBounds(10, 11, 147, 14);
pnlList2.add(lblMZ2);

JLabel lblRT2 = new JLabel("Retention time (min):");
lblRT2.setBounds(150, 11, 147, 14);
pnlList2.add(lblRT2);

JPanel pnlResults = new JPanel();
pnlResults.setLayout(null);
pnlResults.setBorder(new BevelBorder(BevelBorder.LOWERED,
    null, null, null, null));
pnlResults.setBounds(612, 32, 302, 271);
pnlMain.add(pnlResults);

txtResults = new JTextArea();
txtResults.setEditable(false);
txtResults.setBounds(286, 32, 128, 222);

JScrollPane scrollResults = new JScrollPane(txtResults);
scrollResults.setBounds(10, 32, 282, 226);
scrollResults.setVerticalScrollBarPolicy(
    ScrollPaneConstants.VERTICAL_SCROLLBAR_ALWAYS);
```

```
pnlResults.add(scrollResults);

JLabel lblPeakPairs = new JLabel("Peak pairs:");
lblPeakPairs.setBounds(10, 11, 147, 14);
pnlResults.add(lblPeakPairs);

txtDeltamass = new JTextField();
txtDeltamass.setColumns(10);
txtDeltamass.setBounds(106, 316, 90, 20);
pnlMain.add(txtDeltamass);

txtErrorThreshold = new JTextField();
txtErrorThreshold.setText("10");
txtErrorThreshold.setColumns(10);
txtErrorThreshold.setBounds(347, 316, 52, 20);
pnlMain.add(txtErrorThreshold);

btnSearch = new JButton("Search");
btnSearch.addActionListener(this);

txtRTWidth = new JTextField();
txtRTWidth.setText("0.5");
txtRTWidth.setColumns(10);
txtRTWidth.setBounds(552, 316, 52, 20);
pnlMain.add(txtRTWidth);
btnSearch.setBounds(809, 315, 105, 23);
pnlMain.add(btnSearch);

progressBar = new JProgressBar();
progressBar.setStringPainted(true);
progressBar.setString("");
progressBar.setBounds(10, 347, 904, 16);
pnlMain.add(progressBar);

lblSearching = new JLabel("");
lblSearching.setBounds(622, 319, 248, 14);
pnlMain.add(lblSearching);

JLabel lblDeltamass = new JLabel("Delta mass (Da):");
lblDeltamass.setBounds(10, 319, 128, 14);
pnlMain.add(lblDeltamass);

JLabel lblErrorThreshold = new JLabel("Error threshold
    (ppm):");
lblErrorThreshold.setBounds(220, 319, 128, 14);
pnlMain.add(lblErrorThreshold);

JLabel lblRTWidth = new JLabel("RT width (min):  +/-");
lblRTWidth.setBounds(444, 319, 128, 14);
pnlMain.add(lblRTWidth);

JLabel lblResults = new JLabel("Results:");
```

```
        lblResults.setBounds(612, 10, 147, 14);
        pnlMain.add(lblResults);

    }

    /** When an action is performed */

    public void actionPerformed(ActionEvent evt) {
        reset();
        task = new Task();
        task.addPropertyChangeListener(this);
        task.execute();
    }

    /** When task's progress property changes */

    public void propertyChange(PropertyChangeEvent evt) {
        if ("progress" == evt.getPropertyName()) {
            int progress = (Integer) evt.getNewValue();
            progressBar.setValue(progress);
            lblSearching.setText(String.format(
                    "Searching... %d%%\n", task.getProgress()));
        }
    }

    private void reset() {
        txtResults.setText("");
        lblSearching.setText("");
        progressBar.setString("");
        btnSearch.setEnabled(false);
        setCursor(Cursor.getPredefinedCursor(Cursor.WAIT_CURSOR));
    }
}
```

REFERENCES

Abdelmegeed, M.A., Jang, S., Banerjee, A., Hardwick, J.P. and Song, B.J. (2013). Robust protein nitration contributes to acetaminophen-induced mitochondrial dysfunction and acute liver injury. *Free Radical Biology and Medicine*, *60*, 211–222.

Adam, G.C., Sorensen, E.J. and Cravatt, B.F. (2002). Chemical strategies for functional proteomics. *Molecular and Cellular Proteomics*, *1*(10), 781–790.

Aebersold, R. and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, *422*(6928), 198–207.

Ahrends, R., Pieper, S., Kühn, A., Weisshoff, H., Hamester, M., Lindemann, T., Scheler, C., Lehmann, K., Taubner, K. and Linscheid, M.W. (2007). A metal-coded affinity tag approach to quantitative proteomics. *Molecular and Cellular Proteomics*, *6*(11), 1907–1916.

Akiba, S., Kumazawa, S., Yamaguchi, H., Hontani, N., Matsumoto, T., Ikeda, T., Oka, M. and Sato, T. (2006). Acceleration of matrix metalloproteinase-1 production and activation of platelet-derived growth factor receptor beta in human coronary smooth muscle cells by oxidized LDL and 4-hydroxynonenal. *Biochimica et Biophysica Acta*, *1763*(8), 797–804.

Aldini, G., Domingues, M.R., Spickett, C.M., Domingues, P., Altomare, A., Sánchez-Gómez, F.J., Oeste, C.L. and Pérez-Sala, D. (2015). Protein lipoxidation: Detection strategies and challenges. *Redox Biology*, *5*, 253–266.

Alpert, A.J. (2008). Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Analytical Chemistry*, *80*(1), 62–76.

Altelaar, A.F.M. and Heck, A.J.R. (2012). Trends in ultrasensitive proteomics. *Current Opinion in Chemical Biology*, *16*(1-2), 206–213.

Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., Cooke, T.D., Greenwald, R., Hochberg, M., Howell, D., Kaplan, D., Koopman, W., Longley, S., Mankin, H., McShane, D.J., Medsger, T., Meenan, R., Mikkelsen, W., Moskowitz, R., Murphy, W., Rothschild, B., Segal, M., Sokoloff, L. and Wolfe, F. (1986). Development of criteria for the classification and reporting of osteoarthritis: Classification of osteoarthritis of the knee. *Arthritis and Rheumatism*, *29*(8), 1039–1049.

Anderson, N.L. and Anderson, N.G. (1998). Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*, *19*(11), 1853–1861.

Annangudi, S.P., Deng, Y., Gu, X., Zhang, W., Crabb, J.W. and Salomon, R.G. (2008). Low-density lipoprotein has an enormous capacity to bind (E)-4-hydroxynon-2-enal (HNE): detection and characterization of lysyl and histidyl adducts containing multiple molecules of HNE. *Chemical Research in Toxicology*, *21*(7), 1384–1395.

Annesley, T.M. (2003). Ion suppression in mass spectrometry. *Clinical Chemistry*, *49*(7), 1041–1044.

Arnold, R.J., Hrncirova, P., Annaiah, K. and Novotny, M. V. (2004). Fast proteolytic digestion coupled with organelle enrichment for proteomic analysis of rat liver. *Journal of Proteome Research*, *3*(3), 653–657.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I. and Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*, *40*(W1), W597–W603.

Attia, S.M. (2010). Deleterious effects of reactive metabolites. *Oxidative Medicine and Cellular Longevity*, *3*(4), 238–253.

Augusto, F., Hantao, L.W., Mogollón, N.G.S. and Braga, S.C.G.N. (2013). New materials and trends in sorbents for solid-phase extraction. *TrAC - Trends in Analytical Chemistry*, *43*, 14–23.

Aureli, L., Gioia, M., Cerbara, I., Monaco, S., Fasciglione, G., Marini, S., Ascenzi, P., Topai, A. and Coletta, M. (2008). Structural Bases for Substrate and Inhibitor Recognition by Matrix Metalloproteinases. *Current Medicinal Chemistry*, *15*(22), 2192–2222.

Axworthy, D.B., Hoffmann, K.J., Streeter, A.J., Calleman, C.J., Pascoe, G.A. and Baillie, T.A. (1988). Covalent binding of acetaminophen to mouse hemoglobin. Identification of major and minor adducts formed *in vivo* and implications for the nature of the arylating metabolites. *Chemico-Biological Interactions*, *68*(1-2), 99–116.

Badenhorst, C.P.S., van der Sluis, R., Erasmus, E. and van Dijk, A.A. (2013). Glycine conjugation: importance in metabolism, the role of glycine N-

acyltransferase, and factors that influence interindividual variation. *Expert Opinion on Drug Metabolism and Toxicology*, *46*(3), 1–15.

Bakala, H., Ladouce, R., Baraibar, M.A. and Friguet, B. (2013). Differential expression and glycative damage affect specific mitochondrial proteins with aging in rat liver. *Biochimica et Biophysica Acta*, *1832*(12), 2057–2067.

Balbo, S., Hecht, S.S., Upadhyaya, P. and Villalta, P.W. (2014). Application of a High-Resolution Mass-Spectrometry-Based DNA Adductomics Approach for Identification of DNA Adducts in Complex Mixtures. *Analytical Chemistry*, *86*(3), 1744–1752.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C. and Klopp, C. (2014). jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics*, *15*(1), 293.

Barsky, A., Gardy, J.L., Hancock, R.E.W. and Munzner, T. (2007). Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics*, *23*(8), 1040–1042.

Bartolone, J.B., Cohen, S.D. and Khairallah, E.A. (1989). Immunohistochemical localization of acetaminophen-bound liver proteins. *Toxicological Sciences*, *13*(4), 859–862.

Bate, S.T. and Clark, R.A. (2014). *The Design and Statistical Analysis of Animal Experiments*, Cambridge, UK: Cambridge University Press.

Bendz, M., Möller, M.C., Arrigoni, G., Wåhlander, A., Stella, R., Cappadona, S., Levander, F., Hederstedt, L. and James, P. (2009). Quantification of membrane proteins using nonspecific protease digestions. *Journal of Proteome Research*, *8*(12), 5666–5673.

Benson, W.H. and Giulio, R.T. Di (2006). *Genomic Approaches for Cross-Species Extrapolation in Toxicology*, Boca Raton, FL: CRC Press.

Bern, M., Finney, G., Hoopmann, M.R., Merrihew, G., Toth, M.J. and MacCoss, M.J. (2010). Deconvolution of mixture spectra from ion-trap data-independent-acquisition tandem mass spectrometry. *Analytical Chemistry*, *82*(3), 833–841.

Berridge, G., Chalk, R., D'Avanzo, N., Dong, L., Doyle, D., Kim, J.-I., Xia, X., Burgess-Brown, N., Deriso, A., Carpenter, E.P. and Gileadi, O. (2011). High-performance liquid chromatography separation and intact mass analysis of detergent-solubilized integral membrane proteins. *Analytical Biochemistry*, *410*(2), 272–280.

Betancourt, L.H., De Bock, P.-J., Staes, A., Timmerman, E., Perez-Riverol, Y., Sanchez, A., Besada, V., Gonzalez, L.J., Vandekerckhove, J. and Gevaert, K. (2013). SCX charge state selective separation of tryptic peptides combined with 2D-RP-HPLC allows for detailed proteome mapping. *Journal of Proteomics*, *91*, 164–171.

Blackstock, W.P. and Weir, M.P. (1999). Proteomics: Quantitative and physical mapping of cellular proteins. *Trends in Biotechnology*, *17*(3), 121–127.

Blonder, J., Conrads, T.P., Yu, L.-R., Terunuma, A., Janini, G.M., Issaq, H.J., Vogel, J.C. and Veenstra, T.D. (2004). A detergent- and cyanogen bromide-free method for integral membrane proteomics: application to Halobacterium purple membranes and the human epidermal membrane proteome. *Proteomics*, *4*(1), 31–45.

Boichenko, A.P., Govorukhina, N., van der Zee, A.G.J. and Bischoff, R. (2013). Multidimensional separation of tryptic peptides from human serum proteins

using reversed-phase, strong cation exchange, weak anion exchange, and fused-core fluorinated stationary phases. *Journal of Separation Science*, *36*(21-22), 3463–3470.

Bruenner, B.A., Jones, A.D. and German, J.B. (1995). Direct characterization of protein adducts of the lipid peroxidation product 4-hydroxy-2-nonenal using electrospray mass spectrometry. *Chemical Research in Toxicology*, *8*(4), 552–559.

Bucher, J., Riedmaier, S., Schnabel, A., Marcus, K., Vacun, G., Weiss, T.S., Thasler, W.E., Nüssler, A.K., Zanger, U.M. and Reuss, M. (2011). A systems biology approach to dynamic modeling and inter-subject variability of statin pharmacokinetics in human hepatocytes. *BMC Systems Biology*, *5*(1), 66.

Burkhart, J.M., Schumbrutzki, C., Wortelkamp, S., Sickmann, A. and Zahedi, R.P. (2012). Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *Journal of Proteomics*, *75*(4), 1454–1462.

Cao, L., Clifton, J.G., Reutter, W. and Josic, D. (2013). Mass spectrometry-based analysis of rat liver and hepatocellular carcinoma Morris hepatoma 7777 plasma membrane proteome. *Analytical Chemistry*, *85*(17), 8112–8120.

Carpenter, K.J. (1986). The history of enthusiasm for protein. *Journal of Nutrition*, *116*(7), 1364–1370.

Chandramouli, K. and Qian, P.-Y. (2009). Proteomics: Challenges, Techniques and Possibilities to Overcome Biological Sample Complexity. *Human Genomics and Proteomics*, *2009*, 1–22.

Chapel, A., Kieffer-Jaquinod, S., Sagné, C., Verdon, Q., Ivaldi, C., Mellal, M., Thirion, J., Jadot, M., Bruley, C., Garin, J., Gasnier, B. and Journet, A. (2013). An extended proteome map of the lysosomal membrane reveals novel potential transporters. *Molecular and Cellular Proteomics*, *12*(6), 1572–1588.

Chapple, S.J., Cheng, X. and Mann, G.E. (2013). Effects of 4-hydroxynonenal on vascular endothelial and smooth muscle cell redox signaling and function in health and disease. *Redox Biology*, *1*(1), 319–331.

Chen, E.I., Cociorva, D., Norris, J.L. and Yates, J.R. (2007). Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *Journal of Proteome Research*, *6*(7), 2529–2538.

Chen, M., Bisgin, H., Tong, L., Hong, H., Fang, H., Borlak, J. and Tong, W. (2014). Toward predictive models for drug-induced liver injury in humans: are we there yet? *Biomarkers in Medicine*, *8*(2), 201–213.

Chondrogianni, N., Petropoulos, I., Grimm, S., Georgila, K., Catalgol, B., Friguet, B., Grune, T. and Gonos, E.S. (2014). Protein damage, repair and proteolysis. *Molecular Aspects of Medicine*, *35*, 1–71.

Cohen, S.D., Pumford, N.R., Khairallah, E.A., Boekelheide, K., Pohl, L.R., Amouzadeh, H.R. and Hinson, J. a (1997). Selective protein covalent binding and target organ toxicity. *Toxicology and Applied Pharmacology*, *143*(1), 1–12.

Conney, A.H. (2001). In Memoriam: James A. Miller (1915–2000). *Cancer Research*, *61*, 3847–3848.

Copple, I.M., Goldring, C.E., Jenkins, R.E., Chia, A.J.L., Randle, L.E., Hayes, J.D., Kitteringham, N.R. and Park, B.K. (2008). The hepatotoxic metabolite of

acetaminophen directly activates the keap1-Nrf2 cell defense system. *Hepatology*, *48*(4), 1292–1301.

Craig, R. and Beavis, R.C. (2004). TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*, *20*(9), 1466–1467.

Damsten, M.C., Commandeur, J.N.M., Fidder, A., Hulst, A.G., Touw, D., Noort, D. and Vermeulen, N.P.E. (2007). Liquid Chromatography / Tandem Mass Spectrometry Detection of Covalent Binding of Acetaminophen to Human Serum Albumin. *Drug Metabolism and Disposition*, *35*(8), 1408–1417.

De Hoffmann, E. and Stroobant, V. (2007). *Mass Spectrometry* 3rd ed., Chichester, UK: John Wiley & Sons.

Del Villar, K. and Miller, C.A. (2004). Down-regulation of DENN/MADD, a TNF receptor binding protein, correlates with neuronal cell death in Alzheimer's disease brain and hippocampal neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(12), 4210–4215.

De Vos, J., Stassen, C., Vaast, A., Desmet, G. and Eeltink, S. (2012). High-resolution separations of tryptic digest mixtures using core-shell particulate columns operated at 1200bar. *Journal of Chromatography. A*, *1264*, 57–62.

Dietze, E.C., Schäfer, A., Omichinski, J.G. and Nelson, S.D. (1997). Inactivation of glyceraldehyde-3-phosphate dehydrogenase by a reactive metabolite of acetaminophen and mass spectral characterization of an arylated active site peptide. *Chemical Research in Toxicology*, *10*(10), 1097–1103.

Di Palma, S., Boersema, P.J., Heck, A.J.R. and Mohammed, S. (2011). Zwitterionic hydrophilic interaction liquid chromatography (ZIC-HILIC and ZIC-cHILIC)

provide high resolution separation and increase sensitivity in proteome analysis. *Analytical Chemistry*, *83*(9), 3440–3447.

Di Palma, S., Hennrich, M.L., Heck, A.J.R. and Mohammed, S. (2012). Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *Journal of Proteomics*, *75*(13), 3791–3813.

Distler, A.M., Kerner, J., Peterman, S.M. and Hoppel, C.L. (2006). A targeted proteomic approach for the analysis of rat liver mitochondrial outer membrane proteins with extensive sequence coverage. *Analytical Biochemistry*, *356*(1), 18–29.

Donato, P., Cacciola, F., Mondello, L. and Dugo, P. (2011). Comprehensive chromatographic separations in proteomics. *Journal of Chromatography. A*, *1218*(49), 8777–8790.

Doorn, J.A. and Petersen, D.R. (2003). Covalent adduction of nucleophilic amino acids by 4-hydroxynonenal and 4-oxononenal. *Chemico-Biological Interactions*, *143-144*, 93–100.

Druckova, A., Mernaugh, R.L., Ham, A.J.L. and Marnett, L.J. (2007). Identification of the protein targets of the reactive metabolite of teucrin A *in vivo* in the rat. *Chemical Research in Toxicology*, *20*(10), 1393–1408.

Duffus, J.H., Nordberg, M. and Templeton, D.M. (2007). Glossary of terms used in toxicology, 2nd edition (IUPAC Recommendations 2007). *Pure and Applied Chemistry*, *79*(7), 1153–1344.

Dunn, B.M. (2002). Structure and Mechanism of the Pepsin-Like Family of Aspartic Peptidases. *Chemical Reviews*, *102*(12), 4431–4458.

Dunston, C.R., Herbert, R. and Griffiths, H.R. (2015). Improving T cell-induced response to subunit vaccines: opportunities for a proteomic systems approach. *Journal of Pharmacy and Pharmacology*, *67*(3), 290–299.

Edman, P., Högfeldt, E., Sillén, L.G. and Kinell, P.-O. (1950). Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chemica Scandinavica*, *4*, 283–293.

Eichhorn, P., Pérez, S. and Barceló, D. (2012). Time-of-Flight Mass Spectrometry Versus Orbitrap-Based Mass Spectrometry for the Screening and Identification of Drugs and Metabolites. In A. R. Fernandez-Alba, ed. *Comprehensive Analytical Chemistry, Volume 58*. Oxford, UK: Elsevier, pp. 217–272.

Ekman, R., Silberring, J., Westman-Brinkmalm, A.M. and Kraj, A. eds. (2009). *Mass Spectrometry Instrumentation, Interpretation, and Applications*, Hoboken, NJ: John Wiley & Sons.

Eng, J.K., McCormack, A.L. and Yates, J.R. (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, *5*(11), 976–989.

Engel, C.K., Pirard, B., Schimanski, S., Kirsch, R., Habermann, J., Klingler, O., Schlotte, V., Weithmann, K.U. and Wendt, K.U. (2005). Structural Basis for the Highly Selective Inhibition of MMP-13. *Chemistry and Biology*, *12*(2), 181–189.

Esterbauer, H., Schaur, R.J. and Zollner, H. (1991). Chemistry and biochemistry of 4-hydroxynonenal, malonaldehyde and related aldehydes. *Free Radical Biology and Medicine*, *11*(1), 81–128.

Evans, D.C., Watt, A.P., Nicoll-Griffith, D.A. and Baillie, T.A. (2004). Drug–Protein Adducts: An Industry Perspective on Minimizing the Potential for Drug Bioactivation in Drug Discovery and Development. *Chemical Research in Toxicology*, *17*(1), 3–16.

Farrah, T., Deutsch, E.W., Hoopmann, M.R., Hallows, J.L., Sun, Z., Huang, C.-Y. and Moritz, R.L. (2013). The state of the human proteome in 2012 as viewed through PeptideAtlas. *Journal of Proteome Research*, *12*(1), 162–171.

Feist, P. and Hummon, A. (2015). Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples. *International Journal of Molecular Sciences*, *16*(2), 3537–3563.

Fischer, F. and Poetsch, A. (2006). Protein cleavage strategies for an improved analysis of the membrane proteome. *Proteome Science*, *4*(1), 2.

Fontana, R.J. (2014). Pathogenesis of idiosyncratic drug-induced liver injury and clinical perspectives. *Gastroenterology*, *146*(4), 914–928.

Fu, X., Kassim, S.Y., Parks, W.C. and Heinecke, J.W. (2001). Hypochlorous Acid Oxygenates the Cysteine Switch Domain of Pro-matrilysin (MMP-7): A Mechanism for Matrix Metalloproteinase Activation and Atherosclerotic Plaque Rupture by Myeloperoxidase. *Journal of Biological Chemistry*, *276*(44), 41279–41287.

Galeva, N. and Altermann, M. (2002). Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins. *Proteomics*, *2*(6), 713–722.

Galeva, N., Yakovlev, D., Koen, Y., Duzhak, T. and Alterman, M. (2003). Direct Identification of Cytochrome P450 Isozymes by Matrix-Assisted Laser Desorption / Ionization Time of Flight-Based Proteomic Approach. *Drug Metabolism and Disposition*, *31*(4), 351–355.

Gersch, M., Kreuzer, J. and Sieber, S.A. (2012). Electrophilic natural products and their biological targets. *Natural Product Reports*, *29*(6), 659–682.

Gilar, M., Daly, A.E., Kele, M., Neue, U.D. and Gebler, J.C. (2004). Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography. *Journal of Chromatography. A*, *1061*(2), 183–192.

Gilar, M., Olivova, P., Daly, A.E. and Gebler, J.C. (2005). Orthogonality of separation in two-dimensional liquid chromatography. *Analytical Chemistry*, *77*(19), 6426–6434.

Gillet, L.C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R. and Aebersold, R. (2012). Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular and Cellular Proteomics*, *11*(6), O111.016717–O111.016717.

Gilmore, J.M. and Washburn, M.P. (2010). Advances in shotgun proteomics and the analysis of membrane proteomes. *Journal of Proteomics*, *73*(11), 2078–2091.

Golizeh, M., Abusarah, J., Benderdour, M. and Sleno, L. (2014). Covalent Binding of 4-Hydroxynonenal to Matrix Metalloproteinase 13 Studied by Liquid Chromatography–Mass Spectrometry. *Chemical Research in Toxicology*, *27*(9), 1556–1565.

Golizeh, M., LeBlanc, A. and Sleno, L. (2015)(a) Identification of acetaminophen adducts of rat liver microsomal proteins using 2D-LC-MS/MS. *Chemical Research in Toxicology*, *28*(11), 2142–2150.

Golizeh, M., Schneider, C., Ohlund, L.B. and Sleno, L. (2015)(b). Dataset from proteomic analysis of rat, mouse, and human liver microsomes and S9 fractions. *Data in Brief*, *3*, 95–98.

Golizeh, M., Schneider, C., Ohlund, L.B. and Sleno, L. (2015)(c). Multidimensional LC–MS/MS analysis of liver proteins in rat, mouse and human microsomal and S9 fractions. *EuPA Open Proteomics*, *6*, 16–27.

Golizeh, M. and Sleno, L. (2013). Optimized proteomic analysis of rat liver microsomes using dual enzyme digestion with 2D-LC-MS/MS. *Journal of Proteomics*, *82*, 166–178.

Griffin, N.M. and Schnitzer, J.E. (2011). Overcoming key technological challenges in using mass spectrometry for mapping cell surfaces in tissues. *Molecular and Cellular Proteomics*, *10*(2), R110.000935.

Grimsrud, P.A., Xie, H., Griffin, T.J. and Bernlohr, D.A. (2008). Oxidative Stress and Covalent Modification of Protein with Bioactive Aldehydes. *Journal of Biological Chemistry*, *283*(32), 21837–21841.

Gronemeyer, T., Wiese, S., Grinhagens, S., Schollenberger, L., Satyagraha, A., Huber, L.A., Meyer, H.E., Warscheid, B. and Just, W.W. (2013). Localization of Rab proteins to peroxisomes: a proteomics and immunofluorescence study. *FEBS Letters*, *587*(4), 328–338.

Gross, J.H. (2004). *Mass Spectrometry - A Textbook* 1st ed., Heidelberg, Germany: Springer-Verlag.

Gu, Z., Kaul, M., Yan, B., Kridel, S.J., Cui, J., Strongin, A., Smith, J.W., Liddington, R.C. and Lipton, S.A. (2002). S-Nitrosylation of Matrix Metalloproteinases: Signaling Pathway to Neuronal Cell Death. *Science*, *297*(5584), 1186–1190.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H. and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, *17*(10), 994–999.

Han, J., Pope, M., Borchers, C. and M Graves, L. (2002). Mapping of protein phosphorylation by dual enzyme digestion and matrix-assisted laser desorption ionization-quadrupole orthogonal time-of-flight mass spectrometry. *Analytical Biochemistry*, *310*(2), 215–218.

Han, J. and Schey, K.L. (2004). Proteolysis and Mass Spectrometric Analysis of an Integral Membrane: Aquaporin 0. *Journal of Proteome Research*, *3*(4), 807–812.

Hanukoglu, I. (1996). Electron Transfer Proteins of Cytochrome P450 Systems. In *Advances in Molecular and Cell Biology*. pp. 29–56.

Hanzlik, R.P., Fang, J. and Koen, Y.M. (2009). Filling and mining the reactive metabolite target protein database. *Chemico-Biological Interactions*, *179*(1), 38–44.

Hanzlik, R.P., Koen, Y.M., Theertham, B., Dong, Y. and Fang, J. (2007). The reactive metabolite target protein database (TPDB) – a web-accessible resource. *BMC Bioinformatics*, *8*(1), 95.

Hartley, H. (1951). Origin of the Word "Protein." *Nature*, *168*(4267), 244–244.

Havugimana, P.C., Wong, P. and Emili, A. (2007). Improved proteomic discovery by sample pre-fractionation using dual-column ion-exchange high performance liquid chromatography. *Journal of Chromatography. B*, *847*(1), 54–61.

Helbig, A.O., Heck, A.J.R. and Slijper, M. (2010). Exploring the membrane proteome--challenges and analytical strategies. *Journal of Proteomics*, *73*(5), 868–878.

Hennion, M.C. (1999). Solid-phase extraction: method development, sorbents, and coupling with liquid chromatography. *Journal of Chromatography. A*, *856*(1-2), 3–54.

Holčapek, M., Jirásko, R. and Lísa, M. (2012). Recent developments in liquid chromatography-mass spectrometry and related techniques. *Journal of Chromatography. A*, *1259*, 3–15.

House, J.E. (2008). *Inorganic Chemistry*, Oxford, UK: Elsevier.

Hu, X., Duan, Z., Hu, H., Li, G., Yan, S., Wu, J., Wang, J., Yin, D. and Xie, Q. (2013). Proteomic profile of carbonylated proteins in rat liver: exercise attenuated oxidative stress may be involved in fatty liver improvement. *Proteomics*, *13*(10-11), 1755–1764.

Huang, H.-J., Tsai, M.-L., Chen, Y.-W. and Chen, S.-H. (2011). Quantitative shot-gun proteomics and MS-based activity assay for revealing gender differences in enzyme contents for rat liver microsome. *Journal of Proteomics*, *74*(12), 2734–2744.

İçer, M., Zengin, Y., Gunduz, E., Dursun, R., Durgun, H. M., Turkcu, G., Yuksel, H., Üstündağ, M., and Guloglu, C. (2016) Is montelukast as effective as N-

acetylcysteine in hepatic injury due to acetaminophen intoxication in rats? *Experimental and Toxicologic Pathology*, *68*(1), 55–59.

Imming, P., Sinning, C. and Meyer, A. (2006). Drugs, their targets and the nature and number of drug targets. *Nature Reviews. Drug Discovery*, *5*(10), 821–834.

Ioannides, C. ed. (2008). *Cytochromes P450*, Cambridge, UK: Royal Society of Chemistry.

Jaeschke, H., Knight, T.R. and Bajt, M.L. (2003). The role of oxidant stress and reactive nitrogen species in acetaminophen hepatotoxicity. *Toxicology Letters*, *144*(3), 279–288.

Jaeschke, H., McGill, M.R. and Ramachandran, A. (2012). Oxidant stress, mitochondria, and cell death mechanisms in drug-induced liver injury: Lessons learned from acetaminophen hepatotoxicity. *Drug Metabolism Reviews*, *44*(1), 88–106.

Jaeschke, H., McGill, M.R., Williams, C.D. and Ramachandran, A. (2011). Current issues with acetaminophen hepatotoxicity—A clinically relevant model to test the efficacy of natural products. *Life Sciences*, *88*(17-18), 737–745.

Jaeschke, H., Williams, C.D., McGill, M.R., Xie, Y. and Ramachandran, A. (2013). Models of drug-induced liver injury for evaluation of phytotherapeutics and other natural products. *Food and Chemical Toxicology*, *55*, 279–289.

James, L.P., Mayeux, P.R. and Hinson, J.A. (2003). Acetaminophen-induced hepatotoxicity. *Drug Metabolism and Disposition*, *31*(12), 1499–1506.

James, P. (1997). Protein identification in the post-genome era: the rapid rise of proteomics. *Quarterly Reviews of Biophysics*, *30*(4), S0033583597003399.

Jan, Y.H., Heck, D.E., Dragomir, A.C., Gardner, C.R., Laskin, D.L. and Laskin, J.D. (2014). Acetaminophen reactive intermediates target hepatic thioredoxin reductase. *Chemical Research in Toxicology*, *27*(5), 882–894.

Jiang, J., Briedé, J.J., Jennen, D.G.J., Van Summeren, A., Saritas-Brauers, K., Schaart, G., Kleinjans, J.C.S. and de Kok, T.M.C.M. (2015). Increased mitochondrial ROS formation by acetaminophen in human hepatic cells is associated with gene expression changes suggesting disruption of the mitochondrial electron transport chain. *Toxicology Letters*, *234*(2), 139–150.

Jollow, D.J., Mitchell, J.R., Potter, W.Z., Davis, D.C., Gillette, J.R. and Brodie, B.B. (1973). Acetaminophen-induced hepatic necrosis. II. Role of covalent binding *in vivo*. *Journal of Pharmacology and Experimental Therapeutics*, *187*(1), 195–202.

Josephy, P.D. (1997). *Molecular Toxicology* 1st ed., New York, NY: Oxford University Press.

Kalgutkar, A.S. (2011). Handling reactive metabolite positives in drug discovery: What has retrospective structure–toxicity analyses taught us? *Chemico-Biological Interactions*, *192*(1-2), 46–55.

Kalgutkar, A.S., Gardner, I., Obach, R.S., Shaffer, C.L., Callegari, E., Henne, K.R., Mutlib, A.E., Dalvie, D.K., Lee, J.S., Nakai, Y., O'Donnell, J.P., Boer, J. and Harriman, S.P. (2005). A comprehensive listing of bioactivation pathways of organic functional groups. *Current Drug Metabolism*, *6*(3), 161–225.

Kanaeva, I.P., Petushkova, N.A., Lisitsa, A. V, Lokhov, P.G., Zgoda, V.G., Karuzina, I.I. and Archakov, A.I. (2005). Proteomic and biochemical analysis of the mouse liver microsomes. *Toxicology in Vitro*, *19*(6), 805–812.

Kaplowitz, N. (2005). Idiosyncratic drug hepatotoxicity. *Nature Reviews. Drug Discovery*, *4*(6), 489–499.

Kenrick, K.G. and Margolis, J. (1970). Isoelectric focusing and gradient gel electrophoresis: a two-dimensional technique. *Analytical Biochemistry*, *33*(1), 204–207.

Kevorkian, L., Young, D.A., Darrah, C., Donell, S.T., Shepstone, L., Porter, S., Brockbank, S.M. V, Edwards, D.R., Parker, A.E. and Clark, I.M. (2004). Expression profiling of metalloproteinases and their inhibitors in cartilage. *Arthritis and Rheumatism*, *50*(1), 131–141.

Khan, M.F., Bennett, M.J., Jumper, C.C., Percy, A.J., Silva, L.P. and Schriemer, D.C. (2011). Proteomics by mass spectrometry--go big or go home? *Journal of Pharmaceutical and Biomedical Analysis*, *55*(4), 832–841.

Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik*, *26*(3), 231–243.

Kohen, R. and Nyska, A. (2002). Oxidation of Biological Systems: Oxidative Stress Phenomena, Antioxidants, Redox Reactions, and Methods for Their Quantification. *Toxicologic Pathology*, *30*(6), 620–650.

Kong, R.P.W., Siu, S.O., Lee, S.S.M., Lo, C. and Chu, I.K. (2011). Development of online high-/low-pH reversed-phase-reversed-phase two-dimensional liquid chromatography for shotgun proteomics: A reversed-phase-strong cation exchange-reversed-phase approach. *Journal of Chromatography. A*, *1218*(23), 3681–3688.

Koopman, W.J.H., Nijtmans, L.G.J., Dieteren, C.E.J., Roestenberg, P., Valsecchi, F., Smeitink, J.A.M. and Willems, P.H.G.M. (2010). Mammalian mitochondrial complex I: biogenesis, regulation, and reactive oxygen species generation. *Antioxidants and Redox Signaling*, *12*(12), 1431–1470.

Krogh, A., Sonnhammer, E.L.L. and Käll, L. (2007). Advantages of combined transmembrane topology and signal peptide prediction — the Phobius web server. *Nucleic Acids Research*, *35*, 429–432.

Lai, W.G., Zahid, N. and Uetrecht, J.P. (1999). Metabolism of trimethoprim to a reactive iminoquinone methide by activated human neutrophils and hepatic microsomes. *Journal of Pharmacology and Experimental Therapeutics*, *291*(1), 292–299.

Langenfeld, E., Zanger, U.M., Jung, K., Meyer, H.E. and Marcus, K. (2009). Mass spectrometry-based absolute quantification of microsomal cytochrome P450 2D6 in human liver. *Proteomics*, *9*(9), 2313–2323.

Lavén, M., Alsberg, T., Yu, Y., Adolfsson-Erici, M. and Sun, H. (2009). Serial mixed-mode cation- and anion-exchange solid-phase extraction for separation of basic, neutral and acidic pharmaceuticals in wastewater and analysis by high-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Journal of Chromatography. A*, *1216*(1), 49–62.

LeBlanc, A., Shiao, T.C., Roy, R. and Sleno, L. (2015). Development of a sample enrichment protocol using click chemistry for identification of protein targets of reactive metabolites in liver microsomes. In *63rd ASMS Proceedings*. Saint Louis, MO: American Society for Mass Spectrometry (ASMS), p. 1454.

Lee, A.G. (2003). Lipid-protein interactions in biological membranes: A structural perspective. *Biochimica et Biophysica Acta - Biomembranes*, *1612*(1), 1–40.

Lee, S.J., Kim, C.E., Yun, M.R., Seo, K.W., Park, H.M., Yun, J.W., Shin, H.K., Bae, S.S. and Kim, C.D. (2010). 4-Hydroxynonenal enhances MMP-9 production in murine macrophages via 5-lipoxygenase-mediated activation of ERK and p38 MAPK. *Toxicology and Applied Pharmacology*, *242*(2), 191–198.

Lee, S.J., Seo, K.W., Yun, M.R., Bae, S.S., Lee, W.S., Hong, K.W. and Kim, C.D. (2008). 4-Hydroxynonenal enhances MMP-2 production in vascular smooth muscle cells via mitochondrial ROS-mediated activation of the Akt/NF-κB signaling pathways. *Free Radical Biology and Medicine*, *45*(10), 1487–1492.

Le Maire, M., Champeil, P. and Moller, J. V (2000). Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta*, *1508*(1-2), 86–111.

Levivier, E., Goud, B., Souchet, M., Calmels, T.P., Mornon, J.P. and Callebaut, I. (2001). uDENN, DENN, and dDENN: indissociable domains in Rab and MAP kinases signaling pathways. *Biochemical and Biophysical Research Communications*, *287*(3), 688–695.

Li, Q., Jain, M.R., Chen, W. and Li, H. (2013). A multidimensional approach to an in-depth proteomics analysis of transcriptional regulators in neuroblastoma cells. *Journal of Neuroscience Methods*, *216*(2), 118–127.

Liebler, D.C. (2008). Protein Damage by Reactive Electrophiles: Targets and Consequences. *Chemical Research in Toxicology*, *21*(1), 117–128.

Liebler, D.C. and Zimmerman, L.J. (2013). Targeted quantitation of proteins by mass spectrometry. *Biochemistry*, *52*(22), 3797–806.

Lin, Y., Lin, H., Liu, Z., Wang, K. and Yan, Y. (2014). Improvement of a sample preparation method assisted by sodium deoxycholate for mass-spectrometry-based shotgun membrane proteomics. *Journal of Separation Science*, *37*(22), 3321–3329.

Lin, Y., Wang, K., Yan, Y., Lin, H., Peng, B. and Liu, Z. (2013). Evaluation of the combinative application of SDS and sodium deoxycholate to the LC-MS-based shotgun analysis of membrane proteomes. *Journal of Separation Science*, *36*(18), 3026–3034.

Liu, H.Y., Lin, S.L., Chan, S.A., Lin, T.Y. and Fuh, M.R. (2013). Microfluidic chip-based nano-liquid chromatography tandem mass spectrometry for quantification of aflatoxins in peanut products. *Talanta*, *113*, 76–81.

Liu, Z.C., McClelland, R.A. and Uetrecht, J.P. (1995). Oxidation of 5-aminosalicylic acid by hypochlorous acid to a reactive iminoquinone. Possible role in the treatment of inflammatory bowel diseases. *Drug Metabolism and Disposition*, *23*(2), 246–250.

Longworth, J., Noirel, J., Pandhal, J., Wright, P.C. and Vaidyanathan, S. (2012). HILIC- and SCX-based quantitative proteomics of chlamydomonas reinhardtii during nitrogen starvation induced lipid and carbohydrate accumulation. *Journal of Proteome Research*, *11*(12), 5959–5971.

Lourette, N., Smallwood, H., Wu, S., Robinson, E.W., Squier, T.C., Smith, R.D. and Paa-Tolič, L. (2010). A Top-Down LC-FTICR MS-based strategy for characterizing oxidized calmodulin in activated macrophages. *Journal of the American Society for Mass Spectrometry*, *21*(6), 930–939.

Low, T.Y., van Heesch, S., van den Toorn, H., Giansanti, P., Cristobal, A., Toonen, P., Schafer, S., Hübner, N., van Breukelen, B., Mohammed, S., Cuppen, E., Heck, A.J.R. and Guryev, V. (2013). Quantitative and qualitative proteome characteristics extracted from in-depth integrated genomics and proteomics analysis. *Cell Reports*, *5*(5), 1469–1478.

Lynn, D.J., Winsor, G.L., Chan, C., Richard, N., Laird, M.R., Barsky, A., Gardy, J.L., Roche, F.M., Chan, T.H.W., Shah, N., Lo, R., Naseer, M., Que, J., Yau, M., Acab, M., Tulpan, D., Whiteside, M.D., Chikatamarla, A., Mah, B., Munzner, T., Hokamp, K., Hancock, R.E.W. and Brinkman, F.S.L. (2008). InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Molecular Systems Biology*, *4*(218), 218.

Mahmoud, Y., Mahmoud, A., and Nassar, G. (2015) Alpha-lipoic acid treatment of acetaminophen-induced rat liver damage. *Biotechnic and Histochemistry 90*(8), 594–600.

Makarov, A. (2000). Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Analytical Chemistry*, *72*(6), 1156–1162.

Mann, M., Hendrickson, R.C. and Pandey, A. (2001). Analysis of Proteins and Proteomes by Mass Spectrometry. *Annual Review of Biochemistry*, *70*(1), 437–473.

Mann, M., Kulak, N. a, Nagaraj, N. and Cox, J. (2013). The coming age of complete, accurate, and ubiquitous proteomes. *Molecular Cell*, *49*(4), 583–90.

Marnett, L.J. (2000). Oxyradicals and DNA damage. *Carcinogenesis*, *21*(3), 361–370.

Martignoni, M., Groothuis, G.M.M. and de Kanter, R. (2006). Species differences between mouse, rat, dog, monkey and human CYP-mediated drug metabolism, inhibition and induction. *Expert Opinion on Drug Metabolism and Toxicology*, *2*(6), 875–894.

Mast, C., Lyan, B., Joly, C., Centeno, D., Giacomoni, F., Martin, J.-F., Mosoni, L., Dardevet, D., Pujos-Guillot, E., and Papet, I. (2015) Assessment of protein modifications in liver of rats under chronic treatment with paracetamol (acetaminophen) using two complementary mass spectrometry-based metabolomic approaches. *Journal of Proteomics 120*, 194–203.

Mathias, R.A., Chen, Y.-S., Kapp, E.A., Greening, D.W., Mathivanan, S. and Simpson, R.J. (2011). Triton X-114 phase separation in the isolation and purification of mouse liver microsomal membrane proteins. *Methods*, *54*(4), 396–406.

Mazaleuskaya, L.L., Sangkuhl, K., Thorn, C.F., FitzGerald, G.A., Altman, R.B. and Klein, T.E. (2015). PharmGKB summary. *Pharmacogenetics and Genomics*, *25*(8), 416–426.

Mbeunkui, F. and Goshe, M.B. (2011). Investigation of solubilization and digestion methods for microsomal membrane proteome analysis using data-independent LC-MSE. *Proteomics*, *11*(5), 898–911.

McClintock, C.S., Parks, J.M., Bern, M., Ghattyvenkatakrishna, P.K. and Hettich, R.L. (2013). Comparative informatics analysis to evaluate site-specific protein oxidation in multidimensional LC-MS/MS data. *Journal of Proteome Research*, *12*(7), 3307–3316.

McGill, M.R., Williams, C.D., Xie, Y., Ramachandran, A. and Jaeschke, H. (2012). Acetaminophen-induced liver injury in rats and mice: Comparison of protein adducts, mitochondrial dysfunction, and oxidative stress in the mechanism of toxicity. *Toxicology and Applied Pharmacology*, *264*(3), 387–394.

McLafferty, F.W., Breuker, K., Jin, M., Han, X., Infusini, G., Jiang, H., Kong, X. and Begley, T.P. (2007). Top-down MS, a powerful complement to the high capabilities of proteolysis proteomics. *FEBS Journal*, *274*(24), 6256–6268.

Mehmood, S., Allison, T.M. and Robinson, C. V. (2015). Mass Spectrometry of Protein Complexes: From Origins to Applications. *Annual Review of Physical Chemistry*, *66*(1), 453–474.

Michaud, G.A. and Snyder, M. (2002). Proteomic Technologies Review Proteomic Approaches for the Global Analysis of Proteins. *BioTechniques*, *33*(6), 1308–1316.

Michel, P.E., Reymond, F., Arnaud, I.L., Josserand, J., Girault, H.H. and Rossier, J.S. (2003). Protein fractionation in a multicompartment device using Off-Gel isoelectric focusing. *Electrophoresis*, *24*(1-2), 3–11.

Milic, I., Melo, T., Domingues, M.R., Domingues, P. and Fedorova, M. (2015). Heterogeneity of peptide adducts with carbonylated lipid peroxidation products. *Journal of Mass Spectrometry*, *50*(3), 603–612.

Minami, Y., Kawabata, K., Kubo, Y., Arase, S., Hirasaka, K., Nikawa, T., Bando, N., Kawai, Y. and Terao, J. (2009). Peroxidized cholesterol-induced matrix metalloproteinase-9 activation and its suppression by dietary beta-carotene in photoaging of hairless mouse skin. *Journal of Nutritional Biochemistry*, *20*(5), 389–98.

Morquette, B., Shi, Q., Lavigne, P., Ranger, P., Fernandes, J.C. and Benderdour, M. (2006). Production of lipid peroxidation products in osteoarthritic tissues: New evidence linking 4-hydroxynonenal to cartilage degradation. *Arthritis and Rheumatism*, *54*(1), 271–281.

Motoyama, A., Xu, T., Ruse, C.I., Wohlschlegel, J.A. and Yates, J.R. (2007). Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. *Analytical Chemistry*, *79*(10), 3623–3634.

Muldrew, K.L., James, L.P., Coop, L., McCullough, S.S., Hendrickson, H.P., Hinson, J.A. and Mayeux, P.R. (2002). Determination of acetaminophen-protein adducts in mouse liver and serum and human serum after hepatotoxic doses of acetaminophen using high-performance liquid chromatography with electrochemical detection. *Drug Metabolism and Disposition*, *30*(4), 446–451.

Murray, R.K., Bender, D.A., Botham, K.M., Kennelly, P.J., Rodwell, V.W. and Weil, P.A. (2009). Biologic Oxidation. In *Harper's Illustrated Biochemistry, 28th Edition*. McGraw-Hill, p. 704.

Nägele, E., Vollmer, M., Hörth, P. and Vad, C. (2004). 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. *Expert Review of Proteomics*, *1*(1), 37–46.

Nardiello, D., Palermo, C., Natale, A., Quinto, M. and Centonze, D. (2015). Strategies in protein sequencing and characterization: Multi-enzyme digestion coupled with alternate CID/ETD tandem mass spectrometry. *Analytica Chimica Acta*, *854*, 106–117.

Nice, E.C., Rothacker, J., Weinstock, J., Lim, L. and Catimel, B. (2007). Use of multidimensional separation protocols for the purification of trace components

in complex biological samples for proteomics analysis. *Journal of Chromatography. A*, *1168*(1-2), 190–210.

Niemelä, O. (1999). Aldehyde-protein adducts in the liver as a result of ethanol-induced oxidative stress. *Frontiers in Bioscience*, *4*, D506–13.

Nisar, S., Lane, C.S., Wilderspin, A.F., Welham, K.J., Griffiths, W.J. and Patterson, L.H. (2004). A Proteomic Approach to the Identification of Cytochrome P450 Isoforms in Male and Female Rat Liver by Nanoscale Liquid Chromatography-Electrospray Ionization-Tandem Mass Spectrometry. *Drug Metabolism and Disposition*, *32*(4), 382–386.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, *250*(10), 4007–4021.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A. and Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular and Cellular Proteomics*, *4*(10), 1487–1502.

Oliveros, J.C. (2007). VENNY: An interactive tool for comparing lists with Venn Diagrams.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular and Cellular Proteomics*, *1*(5), 376–386.

Overall, C.M. (2000). The Matrix Metalloproteinase (MMP) and Tissue Inhibitor of Metalloproteinase (TIMP) Genes. In I. M. Clark, ed. *Matrix Metalloproteinase Protocols*. Totowa, NJ: Humana Press, pp. 79–120.

Papa, S., Rasmo, D. De, Technikova-Dobrova, Z., Panelli, D., Signorile, A., Scacco, S., Petruzzella, V., Papa, F., Palmisano, G., Gnoni, A., Micelli, L. and Sardanelli, A.M. (2012). Respiratory chain complex I, a main regulatory target of the cAMP/PKA pathway is defective in different human diseases. *FEBS Letters*, *586*(5), 568–577.

Park, B.K., Kitteringham, N.R., Maggs, J.L., Pirmohamed, M. and Williams, D.P. (2005). The Role of Metabolic Activation in Drug-Induced Hepatotoxicity. *Annual Review of Pharmacology and Toxicology*, *45*(1), 177–202.

Parr, R.G. and Pearson, R.G. (1983). Absolute hardness: companion parameter to absolute electronegativity. *Journal of the American Chemical Society*, *105*, 7512–7516.

Peng, F., Zhan, X., Li, M.-Y., Fang, F., Li, G., Li, C., Zhang, P.-F. and Chen, Z. (2012). Proteomic and Bioinformatics Analyses of Mouse Liver Microsomes. *International Journal of Proteomics*, *2012*, 1–24.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J. and Gygi, S.P. (2003). Evaluation of Multidimensional Chromatography Coupled with Tandem Mass Spectrometry (LC/LC–MS/MS) for Large-Scale Protein Analysis: The Yeast Proteome. *Journal of Proteome Research*, *2*(1), 43–50.

Perkins, D.N., Pappin, D.J.C., Creasy, D.M. and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, *20*(18), 3551–3567.

Petushkova, N.A., Kanaeva, I.P., Lisitsa, A. V, Sheremetyeva, G.F., Zgoda, V.G., Samenkova, N.F., Karuzina, I.I. and Archakov, A.I. (2006). Characterization of

human liver cytochromes P450 by combining the biochemical and proteomic approaches. *Toxicology in Vitro*, *20*(6), 966–974.

Piccard, H., Van den Steen, P.E. and Opdenakker, G. (2007). Hemopexin domains as multifunctional liganding modules in matrix metalloproteinases and other proteins. *Journal of Leukocyte Biology*, *81*(4), 870–892.

Plant, N. (2004). Strategies for using *in vitro* screens in drug metabolism. *Drug Discovery Today*, *9*(7), 328–336.

Polettini, A. (2006). *Applications of LC-MS in Toxicology* First., London, UK: Pharmaceutical Press.

Poole, C.F. (2003). New trends in solid-phase extraction. *TrAC - Trends in Analytical Chemistry*, *22*(6), 362–373.

Prabakaran, S., Tepp, W. and Dasgupta, B.R. (2001). Botulinum neurotoxin types B and E: purification, limited proteolysis by endoproteinase Glu-C and pepsin, and comparison of their identified cleaved sites relative to the three-dimensional structure of type A neurotoxin. *Toxicon*, *39*, 1515–1531.

Qiu, Y., Benet, L.Z. and Burlingame, A.L. (1998). Identification of the hepatic protein targets of reactive metabolites of acetaminophen *in vivo* in mice using two-dimensional gel electrophoresis and mass spectrometry. *Journal of Biological Chemistry*, *273*(28), 17940–17953.

Rafiei, A. and Sleno, L. (2015). Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis. *Rapid Communications in Mass Spectrometry*, *29*(1), 119–127.

Reyes, L.H., Encinar, J.R., Marchante-Gayón, J.M., Alonso, J.I.G. and Sanz-Medel, A. (2006). Selenium bioaccessibility assessment in selenized yeast after "*in vitro*" gastrointestinal digestion using two-dimensional chromatography and mass spectrometry. *Journal of Chromatography. A*, *1110*(1-2), 108–116.

Rietschel, B., Arrey, T.N., Meyer, B., Bornemann, S., Schuerken, M., Karas, M. and Poetsch, A. (2009). Elastase digests: new ammunition for shotgun membrane proteomics. *Molecular and Cellular Proteomics*, *8*(5), 1029–1043.

Rietschel, B., Bornemann, S., Arrey, T.N., Baeumlisberger, D., Karas, M. and Meyer, B. (2009). Membrane protein analysis using an improved peptic in-solution digestion protocol. *Proteomics*, *9*(24), 5553–5557.

Roberts, D.W., Bucci, T.J., Benson, R.W., Warbritton, A.R., McRae, T.A., Pumford, N.R. and Hinson, J.A. (1991). Immunohistochemical localization and quantification of the 3-(cystein-S-yl)-acetaminophen protein adduct in acetaminophen hepatotoxicity. *American Journal of Pathology*, *138*(2), 359–371.

Rodríguez, A.A., Ständker, L., Zaharenko, A.J., Garateix, A.G., Forssmann, W.-G., Béress, L., Valdés, O., Hernández, Y. and Laguna, A. (2012). Combining multidimensional liquid chromatography and MALDI-TOF-MS for the fingerprint analysis of secreted peptides from the unexplored sea anemone species Phymanthus crucifer. *Journal of Chromatography. B*, *903*, 30–39.

Roe, M.R., Xie, H., Bandhakavi, S. and Griffin, T.J. (2007). Proteomic Mapping of 4-Hydroxynonenal Protein Modification Sites by Solid-Phase Hydrazide Chemistry and Mass Spectrometry. *Analytical Chemistry*, *79*(10), 3747–3756.

Rombach, E.M. and Hanzlik, R.P. (1999). Detection of Adducts of Bromobenzene 3,4-Oxide with Rat Liver Microsomal Protein Sulfhydryl Groups Using Specific Antibodies. *Chemical Research in Toxicology*, *12*(2), 159–163.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A. and Pappin, D.J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular and Cellular Proteomics*, *3*(12), 1154–1169.

Satterlee, J.S., Basanta-Sanchez, M., Blanco, S., Li, J.B., Meyer, K., Pollock, J., Sadri-Vakili, G. and Rybak-Wolf, a. (2014). Novel RNA Modifications in the Nervous System: Form and Function. *Journal of Neuroscience*, *34*(46), 15170–15177.

Sayre, L.M., Lin, D., Yuan, Q., Zhu, X. and Tang, X. (2006). Protein Adducts Generated from Products of Lipid Oxidation: Focus on HNE and ONE. *Drug Metabolism Reviews*, *38*(4), 651–675.

Schaefer, O., Ohtsuki, S., Kawakami, H., Inoue, T., Liehner, S., Saito, A., Sakamoto, A., Ishiguro, N., Matsumaru, T., Terasaki, T. and Ebner, T. (2012). Absolute quantification and differential expression of drug transporters, cytochrome P450 enzymes, and UDP-glucuronosyltransferases in cultured primary human hepatocytes. *Drug Metabolism and Disposition*, *40*(1), 93–103.

Scheuring, S., Ringler, P., Borgnia, M., Stahlberg, H., Müller, D.J., Agre, P. and Engel, A. (1999). High resolution AFM topographs of the Escherichia coli water channel aquaporin Z. *The EMBO Journal*, *18*(18), 4981–4987.

Seibert, C., Davidson, B.R., Fuller, B.J., Patterson, L.H., Griffiths, W.J. and Wang, Y. (2009). Multiple-approaches to the identification and quantification of cytochromes P450 in human liver tissue by mass spectrometry. *Journal of Proteome Research*, *8*(4), 1672–1681.

Senter, P.D., Al-Abed, Y., Metz, C.N., Benigni, F., Mitchell, R.A., Chesney, J., Han, J., Gartner, C.G., Nelson, S.D., Todaro, G.J. and Bucala, R. (2002). Inhibition of macrophage migration inhibitory factor (MIF) tautomerase and biological activities by acetaminophen metabolites. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(1), 144–149.

Šesták, J., Moravcová, D. and Kahle, V. (2015). Instrument platforms for nano liquid chromatography. *Journal of Chromatography. A*, *1421*, 2–17.

Sharanova, N.E., Kulakova, S.N., Baturina, V.A., Toropygin, I.Y., Khriapova, E.A., Vasilyev, A. V and Gapparov, M.M.G. (2013). Effect of coenzyme Q10 on the proteomic profile of the cytosolic and microsomal fractions from rat hepatocytes upon dietary consumption of various lipid components during ontogeny. *Bulletin of Experimental Biology and Medicine*, *154*(3), 320–325.

Shen, S., Hargus, S.J., Martin, B.M. and Pohl, L.R. (1997). Cytochrome P450 2C11 Is a Target of Diclofenac Covalent Binding in Rats. *Chemical Research in Toxicology*, *10*(4), 420–423.

Shilov, I. V, Seymour, S.L., Patel, A.A., Loboda, A., Tang, W.H., Keating, S.P., Hunter, C.L., Nuwaysir, L.M. and Schaeffer, D.A. (2007). The Paragon Algorithm, a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra. *Molecular and Cellular Proteomics*, *6*(9), 1638–1655.

Shin, N., Liu, Q., Stamer, S.L. and Liebler, D.C. (2007). Protein targets of reactive electrophiles in human liver microsomes. *Chemical Research in Toxicology*, *20*(6), 859–867.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D. and Higgins, D.G. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*(1), 539–539.

Sleno, L., Varesio, E. and Hopfgartner, G. (2007). Determining protein adducts of fipexide: Mass spectrometry based assay for confirming the involvement of its reactive metabolite in covalent binding. *Rapid Communications in Mass Spectrometry*, *21*(24), 4149–4157.

Speers, A.E., Blackler, A.R. and Wu, C.C. (2007). Shotgun analysis of integral membrane proteins facilitated by elevated temperature. *Analytical Chemistry*, *79*(12), 4613–4620.

Speers, A.E. and Wu, C.C. (2007). Proteomics of integral membrane proteins--theory and application. *Chemical Reviews*, *107*(8), 3687–3714.

Streeter, A.J., Dahlin, D.C., Nelson, S.D. and Baillie, T.A. (1984). The covalent binding of acetaminophen to protein. Evidence for cysteine residues as major sites of arylation *in vitro*. *Chemico-Biological Interactions*, *48*(3), 349–366.

Tang, W.H., Shilov, I. V and Seymour, S.L. (2008). Nonlinear fitting method for determining local false discovery rates from decoy database searches. *Journal of Proteome Research*, *7*(9), 3661–3667.

Tao, D., Zhang, L., Shan, Y., Liang, Z. and Zhang, Y. (2011). Recent advances in micro-scale and nano-scale high-performance liquid-phase chromatography for proteome research. *Analytical and Bioanalytical Chemistry*, *399*(1), 229–241.

Thomas, P.D., Kejariwal, A., Campbell, M.J., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S., Vandergriff, J.A. and Doremieux, O. (2003). PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Research*, *31*(1), 334–341.

Titz, B., Elamin, A., Martin, F., Schneider, T., Dijon, S., Ivanov, N. V., Hoeng, J. and Peitsch, M.C. (2014). Proteomics for systems toxicology. *Computational and Structural Biotechnology Journal*, *11*(18), 73–90.

Tousi, F., Bones, J., Iliopoulos, O., Hancock, W.S. and Hincapie, M. (2012). Multidimensional liquid chromatography platform for profiling alterations of clusterin N-glycosylation in the plasma of patients with renal cell carcinoma. *Journal of Chromatography. A*, *1256*, 121–128.

Tran, J.C. and Doucette, A.A. (2008). Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation. *Analytical Chemistry*, *80*(5), 1568–1573.

Tretyakova, N., Villalta, P.W. and Kotapati, S. (2013). Mass Spectrometry of Structurally Modified DNA. *Chemical Reviews*, *113*(4), 2395–2436.

Tsirigos, K.D., Hennerdal, A., Käll, L. and Elofsson, A. (2012). A guideline to proteome-wide α-helical membrane protein topology predictions. *Proteomics*, *12*, 2282–2294.

Tzouros, M. and Pähler, A. (2009). A Targeted Proteomics Approach to the Identification of Peptides Modified by Reactive Metabolites. *Chemical Research in Toxicology*, *22*(5), 853–862.

Uchida, K. and Stadtman, E.R. (1992). Modification of histidine residues in proteins by reaction with 4-hydroxynonenal. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(10), 4544–4548.

Van Montfort, B.A., Doeven, M.K., Canas, B., Veenhoff, L.M., Poolman, B. and Robillard, G.T. (2002). Combined in-gel tryptic digestion and CNBr cleavage for the generation of peptide maps of an integral membrane protein with MALDI-TOF mass spectrometry. *Biochimica et Biophysica Acta*, *1555*(1-3), 111–115.

Van Summeren, A., Renes, J., van Delft, J.H.M., Kleinjans, J.C.S. and Mariman, E.C.M. (2012). Proteomics in the search for mechanisms and biomarkers of drug-induced hepatotoxicity. *Toxicology in Vitro*, *26*(3), 373–385.

Van Swelm, R.P.L., Hadi, M., Laarakkers, C.M.M., Masereeuw, R., Groothuis, G.M.M. and Russel, F.G.M. (2014). Proteomic profiling in incubation medium of mouse, rat and human precision-cut liver slices for biomarker detection regarding acute drug-induced liver injury. *Journal of Applied Toxicology*, *34*(9), 993–1001.

Van Wart, H.E. and Birkedal-Hansen, H. (1990). The cysteine switch: a principle of regulation of metalloproteinase activity with potential applicability to the entire matrix metalloproteinase gene family. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(14), 5578–5582.

Vecchio, P. Del, Graziano, G., Granata, V., Barone, G., Mandrich, L., Rossi, M. and Manco, G. (2002). Denaturing action of urea and guanidine hydrochloride towards two thermophilic esterases. *Biochemical Journal*, *367*(3), 857–863.

Veenstra, T.D. and Smith, R.D. eds. (2003). *Proteome Characterization and Proteomics, Volume 65 of Advances in Protein Chemistry*, San Diego, CA: Academic Press.

Vertommen, A., Panis, B., Swennen, R. and Carpentier, S.C. (2011). Challenges and solutions for the identification of membrane proteins in non-model plants. *Journal of Proteomics*, *74*(8), 1165–1181.

Vizcaíno, J., Deutsch, E. and Wang, R. (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnology*, *32*(3), 223–226.

Vizcaíno, J.A., Côté, R., Reisinger, F., Barsnes, H., Foster, J.M., Rameseder, J., Hermjakob, H. and Martens, L. (2010). The Proteomics Identifications database: 2010 update. *Nucleic Acids Research*, *38*(Database issue), D736–D742.

Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Pérez-Riverol, Y., Reisinger, F., Ríos, D., Wang, R. and Hermjakob, H. (2013). The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Research*, *41*(Database issue), D1063–1069.

Walther, T.C. and Mann, M. (2010). Mass spectrometry-based proteomics in cell biology. *Journal of Cell Biology*, *190*(4), 491–500.

Wang, X., Zhang, A., Wang, P., Sun, H., Wu, G., Sun, W., Lv, H., Jiao, G., Xu, H., Yuan, Y., Liu, L., Zou, D., Wu, Z., Han, Y., Yan, G., Dong, W., Wu, F., Dong, T., Yu, Y., Zhang, S., Wu, X., Tong, X. and Meng, X. (2013). Metabolomics coupled with proteomics advancing drug discovery toward more agile development of targeted combination therapies. *Molecular and Cellular Proteomics*, *12*(5), 1226–1238.

Washburn, M.P., Wolters, D. and Yates, J.R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, *19*(3), 242–247.

Watson, J.T. and Sparkman, O.D. (2007). *Introduction to Mass Spectrometry* 4th ed., Chichester, UK: John Wiley & Sons.

Wendel, A. and Cikryt, P. (1981). Binding of paracetamol metabolites to mouse liver glutathione S-transferases. *Research Communications in Chemical Pathology and Pharmacology*, *33*(3), 463–473.

Westermeier, R. and Naven, T. (2002). *Proteomics in Practice*, Weinheim, Germany: Wiley-VCH.

Williams, D.P. (2006). Toxicophores: investigations in drug safety. *Toxicology*, *226*(1), 1–11.

Wilson, K. and Walker, J.M. eds. (2010). *Principles and Techniques of Biochemistry and Molecular Biology* 7th ed., New York, NY: Cambridge University Press.

Wright, J.C., Beynon, R.J. and Hubbard, S.J. (2010). Cross species proteomics. *Methods in Molecular Biology*, *604*, 123–135.

Wu, Q., Yuan, H., Zhang, L. and Zhang, Y. (2012). Recent advances on multidimensional liquid chromatography-mass spectrometry for proteomics: from qualitative to quantitative analysis--a review. *Analytica Chimica Acta*, *731*, 1–10.

Xie, F., Smith, R.D. and Shen, Y. (2012). Advanced proteomic liquid chromatography. *Journal of Chromatography. A*, *1261*, 78–90.

Xie, Y., Ramachandran, A., Breckenridge, D.G., Liles, J.T., Lebofsky, M., Farhood, A. and Jaeschke, H. (2015). Inhibitor of apoptosis signal-regulating kinase 1 protects against acetaminophen-induced liver injury. *Toxicology and Applied Pharmacology*, 1–9.

Xiong, J. (2006). *Essential Bioinformatics* 1st ed., New York, NY: Cambridge University Press.

Yang, J., Tallman, K.A., Porter, N.A. and Liebler, D.C. (2015). Quantitative Chemoproteomics for Site-Specific Analysis of Protein Alkylation by 4-Hydroxy-2-Nonenal in Cells. *Analytical Chemistry*, *87*(5), 2535–2541.

Yang, Y., Xiao, Q., Humphreys, W.G., Dongre, A. and Shu, Y. (2014). Identification of human liver microsomal proteins adducted by a reactive metabolite using shotgun proteomics. *Chemical Research in Toxicology*, *27*(9), 1537–1546.

Yao, H.-T., Yang, Y.-C., Chang, C.-H., Yang, H.-T., and Yin, M.-C. (2015) Protective effects of (-)-epigallocatechin-3-gallate against acetaminophen-induced liver injury in rats. *BioMedicine 5*(3), 16–21.

Ye, X., Johann, D.J., Hakami, R.M., Xiao, Z., Meng, Z., Ulrich, R.G., Issaq, H.J., Veenstra, T.D. and Blonder, J. (2009). Optimization of protein solubilization for

the analysis of the CD14 human monocyte membrane proteome using LC-MS/MS. *Journal of Proteomics*, *73*(1), 112–122.

Yu, Y., Gilar, M., Lee, P.J., Bouvier, E.S.P. and Gebler, J.C. (2003). Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Analytical Chemistry*, *75*(21), 6023–6028.

Yuan, L. and Kaplowitz, N. (2013). Mechanisms of Drug-induced Liver Injury. *Clinics in Liver Disease*, *17*(4), 507–518.

Yukinaga, H., Iwabuchi, H., Okazaki, O. and Izumi, T. (2012). Glutathione S-transferase pi trapping method for generation and characterization of drug-protein adducts in human liver microsomes using liquid chromatography-tandem mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis*, *67-68*, 186–192.

Zamara, E., Novo, E., Marra, F., Gentilini, A., Romanelli, R.G., Caligiuri, A., Robino, G., Tamagno, E., Aragno, M., Danni, O., Autelli, R., Colombatto, S., Dianzani, M.U., Pinzani, M. and Parola, M. (2004). 4-Hydroxynonenal as a selective pro-fibrogenic stimulus for activated human hepatic stellate cells. *Journal of Hepatology*, *40*(1), 60–68.

Zgoda, V.G., Moshkovskii, S.A., Ponomarenko, E.A., Andreewski, T. V, Kopylov, A.T., Tikhonova, O. V, Melnik, S.A., Lisitsa, A. V and Archakov, A.I. (2009). Proteomics of mouse liver microsomes: performance of different protein separation workflows for LC-MS/MS. *Proteomics*, *9*(16), 4102–4105.

Zhang, A., Sun, H., Wu, G., Sun, W., Yuan, Y. and Wang, X. (2013). Proteomics analysis of hepatoprotective effects for scoparone using MALDI-TOF/TOF mass spectrometry with bioinformatics. *Omics*, *17*(4), 224–229.

Zhang, H., Hou, J., Cui, R., Guo, X., Shi, Z., Yang, F. and Dai, J. (2013). Phosphoproteome analysis reveals an important role for glycogen synthase kinase-3 in perfluorododecanoic acid-induced rat liver toxicity. *Toxicology Letters*, *218*(1), 61–69.

Zhang, L., Yao, L., Zhang, Y., Xue, T., Dai, G., Chen, K., Hu, X. and Xu, L.X. (2012). Protein pre-fractionation with a mixed-bed ion exchange column in 3D LC-MS/MS proteome analysis. *Journal of Chromatography. B*, *905*, 96–104.

Zhang, X. (2015). Less is More: Membrane Protein Digestion Beyond Urea–Trypsin Solution for Next-level Proteomics. *Molecular and Cellular Proteomics*, *14*(9), 2441–2453.

Zhao, Y., Kong, R.P.W., Li, G., Lam, M.P.Y., Law, C.H., Lee, S.M.Y., Lam, H.C. and Chu, I.K. (2012). Fully automatable two-dimensional hydrophilic interaction liquid chromatography-reversed phase liquid chromatography with online tandem mass spectrometry for shotgun proteomics. *Journal of Separation Science*, *35*(14), 1755–1763.

Zhou, L.X., Dehal, S.S., Kupfer, D., Morrell, S., McKenzie, B.A., Eccleston, E.D. and Holtzman, J.L. (1995). Cytochrome P450 catalyzed covalent binding of methoxychlor to rat hepatic, microsomal iodothyronine 5'-monodeiodinase, type I: does exposure to methoxychlor disrupt thyroid hormone metabolism? *Archives of Biochemistry and Biophysics*, *322*(2), 390–394.

Zhou, S. (2003). Separation and detection methods for covalent drug-protein adducts. *Journal of Chromatography. B*, *797*(1-2), 63–90.

Zhou, S., Chan, E., Duan, W., Huang, M. and Chen, Y.-Z. (2005). Drug bioactivation, covalent binding to target proteins and toxicity relevance. *Drug Metabolism Reviews*, *37*(1), 41–213.